



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : G01N 33/574, C07K 16/30, C07H 17/00		A1	(11) International Publication Number: WO 98/05968
			(43) International Publication Date: 12 February 1998 (12.02.98)
(21) International Application Number: PCT/US97/13684			(74) Agents: BAK, Mary, E. et al.; Howson and Howson, Spring House Corporate Center, P.O. Box 457, Spring House, PA 19477 (US).
(22) International Filing Date: 30 July 1997 (30.07.97)			
(30) Priority Data:			
60/022,997 2 August 1996 (02.08.96) US			
60/038,109 19 February 1997 (19.02.97) US			
(60) Parent Applications or Grants			Published <i>With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>
(63) Related by Continuation			
US 60/022,997 (CIP)			
Filed on 2 August 1996 (02.08.96)			
US 60/038,109 (CIP)			
Filed on 19 February 1997 (19.02.97)			
(71) Applicant (for all designated States except US): THE WISTAR INSTITUTE OF ANATOMY AND BIOLOGY [US/US]; 3601 Spruce Street, Philadelphia, PA 19104-4268 (US).			
(72) Inventors; and			
(75) Inventors/Applicants (for US only): RAUSCHER, Frank, J., III. [US/US]; 360 Conestoga Road, Wayne, PA 19087 (US). JENSEN, David, E. [US/US]; 513 South Melville Street, Philadelphia, PA 19143 (US).			
(54) Title: BRCA1 ASSOCIATED PROTEIN (BAP-1) AND USES THEREFOR			
(57) Abstract			
<p>Nucleic acid and amino acid sequences of a BRCA1 Associated Protein, BAP-1, are provided. These sequences, the protein, and anti-BAP-1 antibodies are useful in therapeutics and diagnostics for cancers associated with loss of the 3p2.1 chromosomal region and/or inappropriate BAP-1 levels.</p>			

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

BRCA1 ASSOCIATED PROTEIN (BAP-1) AND USES THEREFOR

This invention was made under work supported by National Institutes of Health, Grant Nos. CA52009, DK49210, and TM54220. The United States Government has
5 certain rights in this invention.

Field of the Invention

This invention relates generally to the field of genes associated with cancers, and particularly, to BRCA1.

10 Background of the Invention

The breast and ovarian cancer susceptibility gene, BRCA1, is linked to the hereditary form of breast cancer. The BRCA1 gene is located on chromosome 17 at the locus 17q21 and encodes a protein of 1863 amino acids. The
15 BRCA 1 locus spans >100 kb comprising 24 exons [Miki et al, Science, 266:66-71 (1994)]. Expression of wild-type BRCA 1 inhibits colony function and tumor growth in vivo, whereas tumor derived mutations of BRCA 1 abolish this growth suppression [Holt et al, Nature Genetics, 12:298-
20 302 (1996)]. Germline mutations in BRCA 1 appear to account for 50% of familial breast cancers and essentially all families with 17q21-linked inherited susceptibility to ovarian and breast cancer [Szabo et al, Hum. Mol. Genet., 4:1811-1817 (1995); Hall et al, Science, 250:1684-1689 (1990); Easton et al, Am J. Hu. Genet., 56:265-271 (1995); Narod et al, Am. J. Hu. Genet., 56:254-264 (1995)]. Kindreds segregating constitutive BRCA1 mutations show a lifetime risk of 40-50% for ovarian cancer and >80% for breast cancer
30 [Easton et al, Am. J. Hum. Genet., 52:678-701 (1993); Easton et al, Am. J. Hum. Genet., 56:265-271 (1995)].

The classification of BRCA 1 as a highly penetrant, autosomal dominant tumor suppressor gene has been

genetically confirmed by the finding of frequent loss or mutation (LOH) of the wild-type allele in breast tumors from mutation carriers [Hall et al, Science, 250:1684-1689 (1990); Miki et al, cited above; Smith et al, Nature Genetics, 2:128-131 (1992)]. Surprisingly, BRCA 1 mutations in sporadic breast cancer including those which show LOH have yet to be found and are extremely rare in sporadic ovarian cancer [Futreal et al, Science, 266:120-122 (1994); Merajver et al, Nature Genetics, 9:439-443 (1995)].

Although the BRCA1 protein resembles no known protein, it does contain a RING domain at its amino terminus [Miki, cited above; Bienstock et al, Cancer Res., 56:2539-2545 (1996)]. The RING finger domain is a complicated structure which chelates two zinc atoms using 7 Cys residues and 1 His residue [C₃HC₄; Lovering et al, Proc. Natl. Acad. Sci. USA, 90:2112-2116 (1993); Freemont et al, Ann. NY Acad. Sci., 684:174-192 (1993)]. This domain is present in a wide variety of proteins with various functions, but the function of the RING finger domain within these proteins is unknown [for a review see Saurin, Trends in Biochem. Sci., 21:208-214 (1996)]. The RING finger of BRCA1 is important to its function since missense mutations in the RING domain (Cys61Gly and Cys64Gly) are found in breast/ovarian kindreds [Friedman et al, Nat. Genet., 8:399-404 (1994); Merajver, cited above; Castilla et al, Nature Genet., 8:387-391 (1994)]. In addition, the RING finger domain is the most conserved region of BRCA1, when comparing the human, mouse and rat proteins. The BRCA1 RING finger is anticipated to be a binding site for protein(s) which either mediate BRCA1 tumor suppressor function or serve to regulate these functions. Genetic evidence supports this in that single amino-acid substitutions at metal chelating cysteines, C61G and C64G, occur in kindreds; these mutations

segregate with the disease susceptibility phenotype and are predicted to abolish RING finger structure.

- Other functions of BRCA1 are discussed in the following references which are incorporated herein by reference:
- 5 Borden et al, EMBO J., 14:1532-1541 (1995);
Lovering et al, Proc. Natl. Acad. Sci. USA, 90:2112-2116
(1993); Koonin et al, Nature Genet., 13:266-268 (1996);
Chen et al, Science, 270:789-791 (1995); Chen et al,
Cancer Research, 56:3168-3172 (1996); Scully et al,
10 Science, 272:123-126 (1996); Thakur et al, Molecular &
Cellular Biology, 17:444-452 (1997); Scully et al, Cell,
88:265-275 (1997); Chapman et al, Nature, 382:678-679
(1996); Scully et al, Proc. Natl. Acad. Sci. USA,
94:5605-5610 (1997); Marquis et al, Nature Genetics,
15 11:17-26 (1995); Gudas et al, Cancer Res., 55:4561-4565
(1995); Gudas et al, Cell Growth and Differentiation,
7:717-723 (1996); Vaughn et al, Cell Growth and
Differentiation, 7:711-715 (1996); Marks et al, Oncogene,
14:115-121 (1997); Zabludoff et al, Oncogene, 13:649-653
20 (1996); Hakem et al, Cell, 85:1009-1023 (1996); Liu et
al, Genes & Development, 10:1835-1843 (1996); Rao et al,
Oncogene, 12:523-528 (1996); Thompson et al, Nature
Genetics, 9:444-450 (1995); Chen et al, J. Biol. Chem.,
271:32863-32868 (1996); Wu et al, Nature Genetics,
25 14:430-440 (1996); Klug et al, FASEB Journal, 9:597-604
(1995); Saurin et al, Trends in Biochem. Sci., 21:208-214
(1996); Friedman et al, Genes & Development, 10:2067-2078
(1996); Neuhausen & Marshall, Cancer Res., 54:6069-6072
(1994); Schildkraut et al, Am. J. Obstet. Gynecol.,
30 172:908-913 (1995); FitzGerald et al, N. Engl. J. Med.,
334:143-149 (1996); Ford et al, Lancet, 343:692-695
(1994); Muto et al, Cancer Research, 56:1250-1252 (1996);
Rao et al, Nature Genetics, 14:185-187 (1996), Struewing
et al, Nature Genetics, 11:198-200 (1995); Couch et al,
35 Human Mutation, 8:8-18 (1996); Holt et al, Nature

Genetics, 12:298-302 (1996); Jensen et al, Nature Genetics, 12:303-308 (1996); Bradley & Sharan, Nature Genetics, 13:268-271 (1996).

There is a need in the art for compositions and methods useful in the treatment and/or prophylaxis of cancers caused by loss of, and mutations in, BRCA1.

Summary of the Invention

The present invention meets the needs in the art by identifying a novel mammalian BRCA1 Associated Protein (BAP-1) and nucleic acid sequences encoding same. BAP1 is the first nuclear-localized ubiquitin carboxy-terminal hydrolase to be identified and is a new tumor suppressor gene which functions in the BRCA1 growth control pathway. Compositions, both diagnostic and therapeutic, based on this newly identified protein are provided herein.

Thus, in one aspect, the present invention provides a nucleic acid sequence, which is isolated from cellular materials with which it is naturally associated. The nucleic acid sequence is preferably selected from SEQ ID NO:1, or a fragment thereof. Such a fragment may have a specified biological function as discussed below, or may encode a peptide having a similar biological function as the intact BAP-1. Homologous nucleotide sequences, and modified nucleotide sequences which encode peptides or proteins which have a similar biological function as the intact BAP-1, are also included in this aspect of the invention.

In another aspect, the present invention provides a mammalian BRCA1 associated protein (BAP-1). In one preferred embodiment, the protein is human and has the amino acid sequence of SEQ ID NO:2. In another embodiment a fragment of the SEQ ID NO:2 encodes a peptide having a similar biological function as the intact BAP-1 protein. Amino acid sequences homologous to

SEQ ID NO: 2, and modified amino acid sequences of SEQ ID NO: 2, which encode peptides or proteins which have a similar biological function as the intact BAP-1 or a specified biological function as discussed below, are also included in this aspect of the invention.

In yet another aspect, the present invention provides a polynucleotide molecule, for example, a vector or plasmid, that comprises a mammalian BAP-1 nucleic acid sequence as defined herein under the control of suitable sequences which direct and regulate expression of the BAP-1 nucleic acid sequence.

In a further aspect, the present invention provides a host cell transformed with a polynucleotide molecule or vector of the invention.

In yet a further aspect, the present invention provides a method of recombinantly expressing BAP-1 or a peptide fragment thereof, by culturing a recombinant host cell according to the invention under conditions which permit expression of BAP-1 or a fragment thereof.

In still a further aspect, the present invention provides an anti-BRCA1 associated protein (BAP-1) antibody.

In yet another aspect, the invention provides a diagnostic reagent comprising an antibody of the invention and a detectable label. Alternatively, a diagnostic reagent of the invention may comprise a nucleic acid sequence of the invention, or a fragment thereof, and a detectable label which is associated with said sequence.

In still another aspect, the invention provides a method of detecting a cancer associated with abnormal levels of BAP-1 comprising providing a biopsy sample from a patient suspected of having said cancer and incubating said sample in the presence of a diagnostic reagent of the invention.

In a further aspect, the present invention provides methods of identifying compounds which specifically bind to BAP-1 or a fragment thereof. In still a further aspect, the present invention provides for compounds or
5 drugs produced by use of the above methods.

Other aspects and advantages of the present invention are described further in the following detailed description of the preferred embodiments thereof.

Brief Description of the Drawings

10 FIG. 1A illustrates the structural features of the BRCA1 gene product. It shows an alignment of RING finger domains of human BRCA1 [SEQ ID NO:3] and mouse BRCA1 [SEQ ID NO:4] (AA1-100), RPT-1 (amino acids 12-100 of SEQ ID NO:5), a putative lymphocyte specific transcription
15 factor having the most closely related RING finger, and BARD1 (AA 47-89) [SEQ ID NO: 19]. Asterisks (*) identify the Zn-chelating amino acids that form the core of the RING finger. Boxed amino acids show regions of identity between the RING finger domains of human BRCA1 and the
20 other proteins. Alignment was performed by ClustalW [Thompson et al, Nucleic Acids Research, 22:4673-4680 (1994)].

FIG. 1B is a schematic map which illustrates the constructs made when the amino terminal 100 amino acids
25 of human BRCA1 (which includes the RING finger domain) and the indicated amino acids of the various BRCA1-RF mutants and controls (described in Example 1) were fused to the LexA DNA-binding domain. The signature C3HC4 structure is highlighted.

30 FIG. 2A provides a comparison of the amino terminal regions of BAP-1 (FLBAP) [amino acids 1-257 of SEQ ID NO: 2], a C. elegans 3 protein [SEQ ID NO: 16], and human ubiquitin carboxyl-terminal hydrolase isozymes L1 (human UBL1) [SEQ ID NO: 17] and L3 (human UBL3) [SEQ ID NO:

18]. Boxed regions indicate areas of greater than 85% homology. This region contains the active sites of UBL1 and UBL3.

FIG. 2B illustrates the sequences and provides a comparison of the partial human [SEQ ID NO: 6] and mouse BAP-1 proteins [SEQ ID NO: 7 - 9] isolated via the yeast 2-hybrid screens of Example 2. Capital letters encode BAP-1. Lower case letters represent the amino acids encoded by the vector. Human BAP-1 is fused to Gal4 activation domain. Mouse BAP-1 is fused to the VP16 activation domain.

FIGS. 3A-3E provide the nucleic acid [SEQ ID NO:1] and amino acid [SEQ ID NO:2] sequences of the novel ubiquitin carboxy-terminal hydrolase, BAP-1. The longest open reading frame which contained the amino acids defined by the (human) 2-hybrid fusion protein is 2188 nucleotides encoding 729 amino acids. The cDNA also contains 39 nucleotides of 5'UTR and 1705 nucleotides of 3'UTR. The enzymatic active site is contained within the first 250 amino acids; the active site residues are circled. The putative nuclear localization signals (NLS) are underlined, the highly acidic region is boxed with heavy lines, the interaction domain is boxed and the protein fragment used to generate BAP1 polyclonal antibodies is bracketed (A.A.'s 483-576 of SEQ ID NO: 2). The conserved amino acids of the ubiquitin COOH-terminal hydrolase active site consensus are circled (amino acids 91, 169, 184 of SEQ ID NO: 2).

FIG. 3F is a comparison of BAP1 (amino acids 1-261 of SEQ ID NO: 2) with other UCH's. UCH-CAEEL (genbank # Q09444) (amino acids 1-251 of SEQ ID NO: 20), UCH DROME (genbank # P35122) [SEQ ID NO: 21] (aa 1-227), YUH1 (genbank # P35127) [SEQ ID NO: 22] (aa 1-236), UCHL-1 (genbank # P09936) [SEQ ID NO: 24] (aa 1-223), UCHL-3 (genbank # P15374) [SEQ ID NO: 23] (aa 1-230). BAP1

(amino acids 630-729 of SEQ ID NO: 2) is further compared to CAEEL-CO8B11.7 (amino acids 238-326 of SEQ ID NO: 20). The BLAST search algorithm was used to identify proteins closely related to BAP1 [Altschul et al, J. Mol. Biol., 215:403-410 (1990)]. The UCH domain of four of these proteins were aligned with BAP1 using the CLUSTALW (ver.1.6) algorithm [Thompson et al, cited above]. Areas of homology with other UCH's are boxed. Only CAEEL-CO8B11.7 showed any homology outside of the enzymatic region.

FIG. 3G is a schematic comparison of the BAP1 and UCH's. The region necessary for the interaction with BRCA1 (AAs 598-729) is indicated in the diagrams with light crosshatching.

15 Detailed Description of the Invention

The present invention provides a novel protein, BRCA1 associated protein-1 (BAP1). BAP1 is a novel, nuclear localized, enzyme which displays the signature motifs and activities of a ubiquitin carboxy-terminal hydrolase, i.e., BAP1 cleaves model ubiquitin substrates in vitro. In fact, BAP1 is the first nuclear-localized ubiquitin carboxy-terminal hydrolase to be identified. The ubiquitin hydrolase function of BAP1 implicates the ubiquitin-proteasome pathway in either the regulation, or as a direct effector, of BRCA1 function. Thus, BAP1 likely has a broad role in ubiquitin-dependent regulatory processes within the nucleus, including the emerging role of ubiquitin conjugation as a subcellular targeting signal, as well as in transcription, chromatin remodeling, cell cycle control and DNA repair/recombination.

BAP1 also enhances the tumor growth suppression properties of BRCA1 in colony formation assays and does so in a manner dependent upon the UCH enzymatic domain

and the BRCA1-interaction domain. BAP1 specifically binds to the wild-type BRCA1 RING finger domain (BRCA1-RF) both in vitro and in vivo, but not to mutant BRCA1-RF's, e.g., the C61G or C64G mutated RING fingers found in tumors from breast cancer kindreds or other closely related RING fingers. The interaction between BAP1 and BRCA1 occurs in vitro and BAP1 mRNA is expressed in those tissues which also express BRCA1. Thus, BAP1 has a role as a tumor suppressor gene.

As described below, the yeast two-hybrid system was employed to isolate mouse and human clones of BAP1. The human BAP1 locus was mapped to human chromosome 3p21.3. Rearrangements and intragenic homozygous deletions and mutations of BAP1 have been found in lung carcinomas, including homozygous deletions found in non-small cell lung cancers.

Together, this evidence supports the role of BAP1 as a tumor suppressor gene and as a regulator or an effector in BRCA1 growth control pathways. Both the specificity of the BRCA1 RING finger-BAP1 interaction and the fact that independent, tumor-derived missense mutations in the cysteines in the BRCA1 RING finger domain abolish interaction with BAP1 provide compelling evidence for the physiological relevance of this interaction.

The invention further provides nucleic acid sequences which encode BAP1 or fragments of BAP1 which have a biological function, diagnostic and therapeutic reagents, as well as methods of using BAP1, its nucleic acid sequences, and antibodies developed thereto. The nucleic acid sequences, protein, amino acid sequences and antibodies directed to BAP1 are useful in the detection, diagnosis and treatment of cancers associated with inappropriate BAP-1 levels and/or loss of chromosomal region 3p21.

In one embodiment, the nucleic acid sequence of the invention is an about 3.5 kb cDNA [SEQ ID NO: 1], encoding BAP-1. BAP-1 is a 729 amino acid protein [SEQ ID NO: 2] which interacts through its carboxy terminus
5 with the BRCA1 RING finger domain. In addition to containing the 250 amino acid amino terminal UCH catalytic domain, it includes a long carboxy-terminal extension with rich in proline, serine and threonine and and contains a short region of extreme acidity in which
10 12 of 13 amino acids are either Glu or Asp, elements which may confer a short half-life upon the protein [Rechsteiner et al, Trends Biochem. Sci., 21:267-271 (1996)]. The extreme carboxy-terminus encodes two potential nuclear localization signals which overlap the
15 approximately 125 amino acid BRCA1-interaction domain. It was this domain that was independently isolated from mouse and human libraries in the two-hybrid screen of Example 2 and is predicted to fold into a long amphipathic helix of coiled-coil character, the structure
20 of which may be important for BRCA1 interaction. Truncation into this region or substitution of a proline for leucine 691 abolish the BAP1-BRCA1 interaction. A potential splice variant in BAP1 results in loss of 31 amino acids of the BRCA1 interaction domain and greatly
25 reduces the ability of BAP1 to bind the BRCA1 RING finger, further suggesting that the BAP1-BRCA1 interaction is regulated. Thus, the BAP1 carboxy-terminus is tethered to BRCA1 via the RING finger domain and that the UCH catalytic domain is free to
30 interact with ubiquitin substrates.

Northern analysis showed that BAP-1 is a ~4kb mRNA expressed in a variety of tissues and cell lines. The cDNA encodes a protein of 80 kD predicted molecular weight. However, expression of the cDNA in vitro or in
35 COS1 cells generated a protein with an apparent molecular

weight of approximately 91 kDa suggesting possible post-translational modifications. Localization of BAP-1 by cell fractionation indicated that it is predominantly a nuclear protein. Chromosomal analysis by fluorescent in situ hybridization (FISH) localized BAP-1 to chromosome 3p21, a genomic region found to be deleted in some breast cancers. The loss of BAP-1 function, individually or in tandem with BRCA1, is anticipated to be associated with breast cancer progression. Thus, the BAP-1 protein may mediate BRCA1 function and inhibit its oncogenic activity. These and other aspects of the invention are discussed in more detail below.

BRCA1 is likely a direct substrate for the UCH activity of BAP1. Thus, in contrast to all of the known UCHs which are comprised entirely of the UCH domain, the carboxy-terminal extension of BAP1 provides substrate and/or targeting specificity for the catalytic function. Regulated ubiquitination of BRCA1 and subsequent proteasome-mediated degradation would not be surprising given that both BRCA1 levels and subnuclear localization are tightly regulated in the mitotic cell-cycle and during meiosis [Gudas et al, cited above; Scully et al, cited above; Zabludoff et al, cited above]. BAP1-mediated deubiquitination of BRCA1 would be expected to stabilize the protein and protect it from proteasome-mediated degradation. This is consistent with both the ability of co-transfected BAP1 to enhance the tumor suppressor effects of BRCA1 in colony formation assays and the finding of mutations in BAP1 in cancer cell lines.

The BRCA1-BAP1 association may also serve to target the UCH domain to other substrates. These substrates may be bound to other sites on BRCA1. BRCA1 could be construed as an assembly or scaffold molecule for regulated assembly of multiprotein complexes, a function

which has been postulated for other tumor suppressor proteins [e.g. pRb; Sellers et al, Biochim. Biophys. Acta., 1288:M1-5 (1996); Welch et al, Genes Dev., 9:31-46 (1995)]. BAP1 may thus be a regulator of this assembly
5 via controlled ubiquitin proteolysis, similar to two other RING finger-containing proteins involved in controlled proteolysis processes, i.e., a mouse homologue of the drosophila seven-in-absentia (sia; a RING finger protein) and the herpes virus protein VMW110 RING finger
10 protein [Everett et al, EMBO J., 16:566-577 (1997)].

The following description defines the aspects of this invention in more detail.

I. Nucleic Acid Sequences

The present invention provides mammalian nucleic
15 acid sequences encoding BAP-1. The nucleic acid sequences of this invention are isolated from cellular materials with which they are naturally associated. In one embodiment, a BAP-1 cDNA sequence is provided in SEQ ID NO:1 (FIGS. 3A-3E).

20 Given the cDNA sequences of SEQ ID NO: 1, one of skill in the art can readily obtain the corresponding anti-sense strands of these DNA sequences. Further, using known techniques, one of skill in the art can readily obtain genomic sequences corresponding to these
25 DNA sequences or the corresponding RNA sequences, as desired.

Similarly the availability of SEQ ID NO: 1 of this invention permits one of skill in the art to obtain other species BAP-1 analogs, by use of the nucleic acid
30 sequences of this invention as probes in a conventional technique, e.g., polymerase chain reaction. Allelic variants of these sequences within a species (i.e., sequences containing some individual nucleotide differences from a more commonly occurring sequence

within a species, but which nevertheless encode the same protein) such as other human variants of BAP-1 SEQ ID NO: 2, may also be readily obtained given the knowledge of this sequence provided by this invention.

5 The present invention further encompasses nucleic acid sequences capable of hybridizing under stringent conditions [see, J. Sambrook et al, Molecular Cloning: A Laboratory Manual, 2d ed., Cold Spring Harbor Laboratory (1989)] to the sequences of SEQ ID NO: 1, their anti-
10 sense strands, or biologically active fragments thereof. An example of a highly stringent hybridization condition is hybridization at 2XSSC at 65°C, followed by a washing in 0.1XSSC at 65°C for an hour. Alternatively, an exemplary highly stringent hybridization condition is in
15 50% formamide, 4XSSC at 42°C. Moderately high stringency conditions may also prove useful, e.g. hybridization in 4XSSC at 55°C, followed by washing in 0.1XSSC at 37°C for an hour. An alternative exemplary moderately high stringency hybridization condition is in 50% formamide,
20 4XSSC at 30°C.

Also encompassed within this invention are fragments of the above-identified nucleic acid sequences. Preferably, such fragments are characterized by encoding a biologically active portion of BAP-1, e.g., an epitope.
25 Generally, these oligonucleotide fragments are at least 15 nucleotides in length. However, oligonucleotide fragments of varying sizes may be selected as desired. Such fragments may be used for such purposes as performing the PCR, e.g., on a biopsied tissue sample.
30 For example, particularly useful fragments of BAP-1 cDNA [SEQ ID NO:1] and corresponding sequences include the open reading frame, nt 40-2226, the nuclear localization sites, nt 2005 to 2022 and nt 2188 to 2205, a region of acidity at nt 1225 to 1263, and the BRCA1-RF-interactive
35 domain at nt 1831 to 2226 of SEQ ID NO:1. Other

fragments which are contained within the above identified fragments or which overlap them and demonstrate similar biological activities, e.g., those which differ by 1 to 9 bases, are also desirable. Similarly, other useful fragments may be readily identified by one of skill in the art by resort to conventional techniques, such as, by deletion mutagenesis, fusion to other proteins, or by motif searches in computer databases. In addition, other suitable techniques are known.

10 The nucleotide sequences of the invention may be isolated by conventional uses of polymerase chain reaction or cloning techniques such as those described in obtaining the murine and human sequences, described below. Alternatively, these sequences may be constructed using conventional genetic engineering or chemical synthesis techniques.

20 According to the invention, the nucleic acid sequences may be modified. Utilizing the sequence data in FIGS. 3A-3E [SEQ ID NO: 1] and in the sequence listing, it is within the skill of the art to obtain other polynucleotide sequences encoding the proteins of the invention. Such modifications at the nucleic acid level include, for example, modifications to the nucleotide sequences which are silent or which change the amino acids, e.g. to improve expression or secretion. Also included are allelic variations, caused by the natural degeneracy of the genetic code.

25 Also encompassed by the present invention are mutants of the BAP-1 gene provided herein. Such mutants include amino terminal, carboxy terminal or internal deletions which are useful as dominant inhibitor genes. Such a truncated, or deletion, mutant may be expressed for the purpose of inhibiting the activity of the full-length or wild-type gene.

These nucleic acid sequences are useful for a variety of diagnostic and therapeutic uses. Advantageously, the nucleic acid sequences are useful in the development of diagnostic probes and antisense probes for use in the detection and diagnosis of conditions characterized by BRCA1 mutation. Additionally, the BAP-1 gene has been mapped to chromosome 3p21.3. Thus, these sequences provide a good marker for further analysis of chromosome 3. The nucleic acid sequences of this invention are also useful in the production of mammalian, and particularly, human BAP-1 proteins and peptides.

II. Protein Sequences

The present invention also provides mammalian BAP-1 polypeptides, peptides or proteins. These proteins are free from association with other contaminating proteins or materials with which they are found in nature. In one embodiment, the invention provides a human BAP-1 [SEQ ID NO:2] polypeptide of 729 amino acids having a predicted molecular weight (MW) of about 81 kD. In another embodiment, the invention provides partial human and murine BAP-1 proteins [SEQ ID NO: 6-9] (FIG. 2B).

Also included in the invention are analogs, or modified versions, of the proteins provided herein. Typically, such analogs differ by only one to four codon changes. Examples include polypeptides with minor amino acid variations from the illustrated amino acid sequences of BAP-1 (FIGS. 3A-3E) [SEQ ID NO: 2]; in particular, conservative amino acid replacements. Conservative replacements are those that take place within a family of amino acids that are related in their side chains and chemical properties. Also provided are homologs of the proteins of the invention which are characterized by having at least about 85% or higher homology with SEQ ID

NO:2. Based on the sequence information provided herein, one of skill in the art can readily obtain BAP-1 from other mammalian species.

Further encompassed by this invention are fragments of the BAP-1 polypeptide. Such fragments are desirably characterized by having BAP-1 biological activity, including, e.g., the ability to bind specifically to the RING finger domain of wild-type BRCA1. These fragments may be designed or obtained in any desired length, including as small as about 5-8 amino acids in length. Such a fragment may represent an epitope of the protein. Alternatively, the BAP-1 proteins [SEQ ID NO:2] of the invention may be modified, for example, by truncation at the amino or carboxy termini, by elimination or substitution of one or more amino acids, or by any number of now conventional techniques to improve production thereof, to enhance protein stability or other characteristics, e.g. binding activity or bioavailability, or to confer some other desired property upon the protein.

Currently, desirable proteins or peptides correspond to the nuclear localization sites, residues 656 to 661 and residues 717 to 722 of SEQ ID NO:2, a region of extreme acidity, residues 396 to 408 SEQ ID NO:2, and the interactive domain, residues 598 to 729 of SEQ ID NO:2. Another suitable fragment, which has homology to ubiquitin carboxyl-terminal hydrolase, isozyme L3, is located between about amino acids 1 to about 214 of SEQ ID NO: 2. Yet another suitable fragment, corresponding to residues 483 to 576 of SEQ ID NO:2 has been used to generate antibodies. Other suitable fragments include amino acids 1 to 313, 1 to 325, 1 to 352, and 1 to 426 of SEQ ID NO: 2. Additionally, fragments which are about in the range of the above amino acid residues, e.g., which differ by 1 to 5 amino acids, are anticipated to be

particularly desirable. Still other suitable BAP-1 fragments are identified in the Examples or may be readily identified and prepared by one of skill in the art using known techniques, such as deletion mutagenesis and expression.

III. Expression

A. In Vitro

To produce recombinant BAP-1 proteins of this invention, the DNA sequences of the invention are inserted into a suitable expression system. Desirably, a recombinant molecule or vector is constructed in which the polynucleotide sequence encoding BAP-1 is operably linked to a heterologous expression control sequence permitting expression of the BAP-1 protein. Numerous types of appropriate expression vectors are known in the art for mammalian (including human) protein expression, by standard molecular biology techniques. Such vectors may be selected from among conventional vector types including insects, e.g., baculovirus expression, or yeast, fungal, bacterial or viral expression systems. Other appropriate expression vectors, of which numerous types are known in the art, can also be used for this purpose.

Methods for obtaining such expression vectors are well-known. See, Sambrook et al, Molecular Cloning. A Laboratory Manual, 2d edition, Cold Spring Harbor Laboratory, New York (1989); Miller et al, Genetic Engineering, 8:277-298 (Plenum Press 1986) and references cited therein.

Suitable host cells or cell lines for transfection by this method include mammalian cells, such as Human 293 cells, Chinese hamster ovary cells (CHO), the monkey COS-1 cell line or murine 3T3 cells derived from Swiss, Balb-c or NIH mice may be used. Another

suitable mammalian cell line is the CV-1 cell line. Still other suitable mammalian host cells, as well as methods for transfection, culture, amplification, screening, production, and purification are known in the art. [See, e.g., Gething and Sambrook, Nature, 293:620-625 (1981), or alternatively, Kaufman et al, Mol. Cell. Biol., 5(7):1750-1759 (1985) or Howley et al, U. S. Patent 4,419,446].

Similarly bacterial cells are useful as host cells for the present invention. For example, the various strains of E. coli (e.g., HB101, MC1061, and strains used in the following examples) are well-known as host cells in the field of biotechnology. Various strains of B. subtilis, Pseudomonas, other bacilli and the like may also be employed in this method.

Many strains of yeast cells known to those skilled in the art are also available as host cells for expression of the polypeptides of the present invention. Other fungal cells may also be employed as expression systems.

Alternatively, insect cells such as Spodoptera frugipedera (Sf9) cells may be used.

Thus, the present invention provides a method for producing a recombinant BAP-1 protein which involves transfecting a host cell with at least one expression vector containing a polynucleotide of the invention under the control of a transcriptional regulatory sequence, e.g., by conventional means such as electroporation. The transfected or transformed host cell is then cultured under conditions that allow expression of the BAP-1 protein. The expressed protein may then be recovered, isolated, and optionally purified from the cell (or from the culture medium, if expressed extracellularly) by appropriate means known to one of skill in the art.

For example, the proteins may be isolated in soluble form following cell lysis, or may be extracted using known techniques, e.g., in guanidine chloride. If desired, the BAP-1 proteins of the invention may be produced as a fusion protein. For example, it may be desirable to produce BAP-1 fusion proteins, to enhance expression of the protein in a selected host cell, to improve purification, or for use in monitoring the presence of BAP-1 in tissues, cells or cell extracts. Suitable fusion partners for the BAP-1 proteins of the invention are well known to those of skill in the art and include, among others, β -galactosidase, glutathione-S-transferase, and poly-histidine.

B. In Vivo

Alternatively, where it is desired that the BAP-1 protein (whether full-length or a desirable fragment) be expressed in vivo, e.g., for gene therapy purposes, an appropriate vector for delivery may be readily selected by one of skill in the art. Exemplary gene therapy vectors are readily available from a variety of academic and commercial sources, and include, e.g., adeno-associated virus [International patent application No. PCT/US91/03440], adenovirus vectors [M. Kay et al, Proc. Natl. Acad. Sci. USA, 91:2353 (1994); S. Ishibashi et al, J. Clin. Invest., 92:883 (1993)], or other viral vectors, e.g., various poxviruses, vaccinia, etc. Methods for insertion of a desired gene, e.g., BAP-1, and obtaining in vivo expression of the encoded protein, are well known to those of skill in the art.

30 IV. Antisera and Antibodies

The BAP-1 proteins of this invention are also useful as antigens for the development of anti-BAP-1 antisera and antibodies to BAP-1 or to a desired fragment of a BAP-1 protein. Specific antisera may be generated using

known techniques. See, Sambrook, cited above, Chapter 18, generally, incorporated by reference and Example 5 below. Similarly, antibodies of the invention, both polyclonal and monoclonal, may be produced by

5 conventional methods. These techniques may include the Kohler and Milstein hybridoma technique, recombinant techniques, such as described by Huse et al, Science, 246:1275-1281 (1988), or any other techniques known to the art.

10 Also encompassed within this invention are humanized and chimeric antibodies. As used herein, a humanized antibody is defined as an antibody containing murine complementary determining regions (CDRs) capable of binding to BAP-1 or a fragment thereof, and human
15 framework regions. These CDRs are preferably derived from a murine monoclonal antibody (MAb) of the invention. As defined herein, a chimeric antibody is defined as an antibody containing the variable region light and heavy chains, including both CDR and framework regions, from a
20 BAP-1 MAb of the invention and the constant region light and heavy chains from a human antibody. Methods of identifying suitable human framework regions and modifying a MAb of the invention to contain same to produce a humanized or chimeric antibody of the
25 invention, are well known to those of skill in the art. See, e.g., E. Mark and Padlin, "Humanization of Monoclonal Antibodies", Chapter 4, The Handbook of Experimental Pharmacology, Vol. 113, The Pharmacology of Monoclonal Antibodies, Springer-Verlag (June, 1994).
30 Other types of recombinantly-designed antibodies are also encompassed by this invention.

Further provided by the present invention are anti-idiotypic antibodies (Ab2) and anti-anti-idiotypic antibodies (Ab3). Ab2 are specific for the target to
35 which anti-BAP-1 antibodies of the invention bind and Ab3

are similar to BAP-1 antibodies (Ab1) in their binding specificities and biological activities [see, e.g., M. Wettendorff et al., "Modulation of anti-tumor immunity by anti-idiotypic antibodies." In Idiotypic Network and Diseases, ed. by J. Cerny and J. Hiernaux J, Am. Soc. Microbiol., Washington DC: pp. 203-229, (1990)]. These anti-idiotypic and anti-anti-idiotypic antibodies may be produced using techniques well known to those of skill in the art. Such anti-idiotypic antibodies (Ab2) can bear the internal image of BAP-1 and bind to BRCA1 in much the same manner as BAP-1, and are thus useful for the same purposes as BAP-1.

In general, polyclonal antisera, monoclonal antibodies and other antibodies which bind to BAP-1 as the antigen (Ab1) are useful to identify epitopes of BAP-1, to separate BAP-1 from contaminants in living tissue (e.g., in chromatographic columns and the like), and in general as research tools and as starting material essential for the development of other types of antibodies described above. Anti-idiotypic antibodies (Ab2) are useful for binding BRCA1 and thus may be used in the treatment of cancers. The Ab3 antibodies may be useful for the same reason the Ab1 are useful. Other uses as research tools and as components for separation of BAP-1 from other contaminant of living tissue, for example, are also contemplated for the above-described antibodies.

V. Diagnostic Reagents and Methods

Advantageously, the present invention provides reagents and methods useful in detecting and diagnosing abnormal levels of BAP-1, (i.e., deficiencies or excesses thereof) in a patient. Conditions associated with excess levels of BAP-1 may be indicative of BRCA1 mutations. Abnormal levels of BAP-1 may be associated with a variety

of cancers, including lung cancer (small cell lung carcinoma and non-small cell lung carcinoma), breast cancers, uterine carcinomas, and oral squamous cell carcinomas, among others.

5 Thus, the proteins, protein fragments, antibodies, and polynucleotide sequences (including anti-sense polynucleotide sequences and oligonucleotide fragments), and BAP-1 antisera and antibodies of this invention may be useful as diagnostic reagents. These reagents may
10 optionally be labelled using diagnostic labels, such as radioactive labels, colorimetric enzyme label systems and the like conventionally used in diagnostic or therapeutic methods. Alternatively, the N- or C-terminus of BAP-1 or a fragment thereof may be tagged with a viral epitope
15 which can be recognized by a specific antisera. The reagents may be used to measure abnormal BAP-1 levels in selected mammalian tissue using conventional diagnostic assays, e.g., Southern blotting, Northern and Western blotting, polymerase chain reaction (PCR), reverse
20 transcriptase (RT) PCR, immunostaining, and the like. For example, in biopsies of tumor tissue, loss of BAP-1 expression in tumor tissue could be directly verified by RT-PCR or immunostaining. Alternatively, a Southern analysis, genomic PCR, or fluorescence in situ
25 hybridization (FISH) may be performed to confirm BAP-1 gene rearrangement.

 In one example, as diagnostic agents the polynucleotide sequences may be employed to detect or quantitate normal BAP-1. The selection of the
30 appropriate assay format and label system is within the skill of the art and may readily be chosen without requiring additional explanation by resort to the wealth of art in the diagnostic area.

 Thus the present invention provides methods for the
35 detection of disorders characterized by inappropriate

BAP-1 levels. The protein, antibody, antisera and polynucleotide reagents of the invention are expected to be useful in the following methods. The methods involve contacting a selected mammalian tissue, e.g., a biopsy sample or other cells, with the selected reagent, protein, antisera antibody or DNA sequence, and measuring or detecting the amount of BAP-1 present in the tissue in a selected assay format based on the binding or hybridization of the reagent to the tissue.

10 VI. Therapeutic Compositions and Methods

BAP-1 is believed to have a role in modulating the activity of BRCA1, a tumor suppressor. More particularly, BAP-1 enzymatic activity is anticipated to have a role in the persistence of BRCA1 in a cell. For example, the extended presence of BRCA1, particularly in high levels, is associated with cell death. Thus, by adjusting BAP-1 levels in a cell, e.g., by use of BAP-1 or an inhibitor identified by the invention, persistence of BRCA-1 in the cells can thereby be altered. For example, it may be desirable to adjust BAP-1 levels so as to enhance BRCA1 persistence in a cell, e.g., a tumor cell. Alternatively, it may be desirable to adjust BAP-1 levels so as to increase BRCA1 degradation in the cell. The compositions and methods useful for the treatment of conditions associated with inadequate or undesirable BAP-1 levels are provided. As stated above, included among such conditions are liver and breast cancers.

The therapeutic compositions of the invention may be formulated to contain an anti-idiotypic antibody of the invention, the BAP-1 protein itself or a fragment thereof. The therapeutic composition desirably contains 0.01 μ g to 10 mg protein. These compositions may contain a pharmaceutically acceptable carrier. Suitable carriers

are well known to those of skill in the art and include, for example, saline. Alternatively, such compositions may include conventional delivery systems into which protein of the invention is incorporated. Optionally,
5 these compositions may contain other active ingredients, e.g., chemotherapeutics.

Still another method involves the use of the BAP-1 polynucleotide sequences for gene therapy. In the method, the BAP-1 sequences are introduced into a
10 suitable vector for delivery to a cell containing a deficiency of BAP-1 and/or to block tumor growth. By conventional genetic engineering techniques, the BAP-1 gene sequence may be introduced to mutate the existing gene by recombination or to replace an inactive or
15 missing gene.

Generally, a suitable vector-based treatment contains between 1×10^{-3} pfu to 1×10^{12} pfu per dose. However, the dose, timing and mode of administration of these compositions may be determined by one of skill in
20 the art. Such factors as the age, condition, and the level of the BAP-1 deficiency detected by the diagnostic methods described above, may be taken into account in determining the dose, timing and mode of administration of the therapeutic compositions of the invention.
25 Generally, where treatment of an existing cancer is indicated, a therapeutic composition of the invention is preferably administered in a site-directed manner and is repeated as needed. Such therapy may be administered in conjunction with conventional therapies, including
30 radiation and/or chemotherapeutic treatments.

VII. Drug Screening and Development

The proteins, antibodies and polynucleotide sequences of the present invention may also be used in the screening and development of chemical compounds or

proteins which have utility as therapeutic drugs for the treatment of cancers characterized by BAP-1 and/or BRCA1 mutation or loss. As one example, a compound capable of binding to BAP-1 and preventing its biological activity
5 may be a useful drug component for the treatment or prevention of cancer. The methods described herein may also be applied to fragments of BAP-1.

Suitable assay methods may be readily determined by one of skill in the art. Where desired, and depending on
10 the assay selected, BAP-1 may be immobilized directly or indirectly (e.g., via an anti-BAP-1 antibody) on a suitable surface, e.g., in an ELISA format. Such immobilization surfaces are well known. For example, a wettable inert bead may be used. Alternatively, BAP-1
15 may be used in screening assays which do not require immobilization, e.g., in the screening of combinatorial libraries.

Assays and techniques exist for the screening and development of drugs capable of binding to selected
20 regions of BAP-1. These include the use of phage display system for expressing the BAP-1 proteins, and using a culture of transfected E. coli or other microorganism to produce the proteins for binding studies of potential binding compounds. See, for example, the techniques
25 described in G. Cesarini, FEBS Letters, 307(1):66-70 (July 1992); H. Gram et al., J. Immunol. Meth., 161:169-176 (1993); C. Summer et al., Proc. Natl. Acad. Sci., USA, 89:3756-3760 (May 1992), incorporated by reference herein.

Other conventional drug screening techniques may be employed using the proteins, antibodies or polynucleotide sequences of this invention. As one example, a method for identifying compounds which specifically bind to a BAP-1 protein can include simply the steps of contacting
30 a selected BAP-1 protein with a test compound to permit
35

binding of the test compound to BAP-1; and determining the amount of test compound, if any, which is bound to the BAP-1 protein. Such a method may involve the incubation of the test compound and the BAP-1 protein immobilized on a solid support.

Typically, the surface containing the immobilized ligand is permitted to come into contact with a solution containing the BAP-1 protein and binding is measured using an appropriate detection system. Suitable detection systems include the streptavidin horse radish peroxidase conjugate, direct conjugation by a tag, e.g., fluorescein. Other systems are well known to those of skill in the art. This invention is not limited by the detection system used.

Another method of identifying compounds which specifically bind to BAP-1 can include the steps of contacting a BAP-1 protein immobilized on a solid support with both a test compound and the protein sequence which is a receptor for BAP-1 to permit binding of the receptor to the BAP-1 protein; and determining the amount of the receptor which is bound to the BAP-1 protein. The inhibition of binding of the normal protein by the test compound thereby indicates binding of the test compound to the BAP-1 protein.

Thus, through use of such methods, the present invention is anticipated to provide compounds capable of interacting with BAP-1 or portions thereof, and either enhancing or decreasing its biological activity, as desired. Such compounds are believed to be encompassed by this invention.

The assay methods described herein are also useful in screening for inhibition of the interaction between a BAP-1 protein of the invention and its ligand(s). The solution containing the inhibitors may be obtained from any appropriate source, including, for example, extracts

of supernatants from culture of bioorganisms, extracts from organisms collected from natural sources, chemical compounds, and mixtures thereof.

These examples illustrate the preferred methods for obtaining and using the sequences and compositions of the invention. These examples are illustrative only and do not limit the scope of the invention.

Example 1 - Construction of Expression Plasmids

A. LexA Fusions

10 A totally synthetic BRCA1 gene encoding the amino-terminal 100 amino acids of human BRCA1 (BRCA1-RF), including the full ring-finger domain, was constructed. The BRCA1-RF domain was made using long overlapping oligonucleotides and PCR-mediated overlap-extension gene
15 synthesis techniques [Madden et al, Science, 253:1550-1553 (1991)]. Codon usage was optimized for expression in *E. coli* and *S. cerevisiae* [Sharp et al, Nuc. Acids Res., 16:8207-8211 (1988)] (Figs. 1A and 1B). The following oligonucleotides were used.

20 top strand: [SEQ ID NO: 10]

5'-ATGGAACCTGTCTGCTCTGCGTGTGAAGAAGTTCAAAACGTTATCAACGCTA-TGCAAAAGATCCTGGAATGTCCAATCTG

bottom strand: [SEQ ID NO: 11]

5'-GGTTCAGCAGCTTCAGCATAACAGAACTTACAGAAGATGTGGTCACACTTAGTG-
25 GAAACTGGTTCCTTGATCAGTTCCAGACAGATTGGACATTCCAGGATC

top strand: [SEQ ID NO: 12]

5'-GTATGCTGAAGCTGCTGAACCAAAAGAAGGGTCCATCTCAATGTCCACTGTG-TAAGAACGACATCACTAAGCGTTCTCTGCAAGAATCTACTCGTTTCTCTC

bottom strand: [SEQ ID NO: 13]

30 5'-TTCCAGACCAGTGTCCAGCTGGAAAGCACAGATGATCTTCAGCAGTTCTTCA-ACCAGTTGAGAGAAACGAGTAGATTCTTG

Double-stranded DNA was generated by 5 cycles of the polymerase chain reaction (PCR) and the full-length cDNA was amplified further via PCR using "outside"

primers with homology to the 5' and 3' ends of the DNA sequence [Madden, cited above]. These primers contained enzymatic restriction sites for either EcoRI (BRCA1-RF-5' primer): 5'-GCTAGAATTACCATGGACCTGTCTCCTCTG [SEQ ID NO: 14] or Sal I (BRCA1-RF-3' primer): 5'-GCTAGTCGACTTCCAGACCAGTGTCCAG [SEQ ID NO: 15].

The resulting complete, "wild-type" RF domain was confirmed by sequencing. The resulting BRCA1-RF was then fused in frame with the LexA DNA-binding domain to create a LexA-BRCA1-RF fusion construct by cloning the BRCA1-RF domain into the EcoRI-SalI restriction sites of the vector pBTM-116 [Vojtek et al, *Cell*, 74:205-214 (1993)] (see Fig. 1B). This LexA-BRCA1-RF construct was used as the probe ("bait") to screen for BRCA1-interacting proteins in a yeast 2-hybrid analysis.

Negative control/specificity controls for the specificity of the interaction in the yeast system were made (as LexA fusions) by mutating the BRCA1-RF (Figs. 1A and 1B) as follows:

(i) The Cys61Gly and Cys64Gly substitutions of BRCA1 which occur in breast cancer pedigrees. BRCA1 RF domain point mutants, BRCA1-C64G (Cys 64 to Gly) and BRCA1-C61 G (Cys 61 to Gly), were created by PCR-mutagenesis using the "outside" primers described above and overlapping oligonucleotides containing the appropriate nucleotide change: BRCA1-C61 G-sense: 5'-CCATCTCAAGGTCCACTGTGTAAG-3' [SEQ ID NO: 25]; BRCA1-C61 G-antisense: 5'-CTTACACAGTGGACCTTGAGATGG-3' [SEQ ID NO: 26]; BRCA1-C64G-sense: 5'-CAATGTCCACTGGGTAAGAACGACATC-3' [SEQ ID NO: 27]; and BRCA1-C64G-antisense: 5'-GATGTCGTTCTTACCCAGTGGACATTG-3' [SEQ ID NO: 28] [Ho et al, *Gene*, 77:51-59 (1989)]. The BRCA1(C64G)-RF control has a point mutation in the BRCA1-RF found in a breast cancer kindred [Castilla et al, *Nat. Genet.*, 8:387-391 (1994)]. This mutation, a

Cys64 to Gly64, destroys one of the Zn chelating residues leading, presumably, to the loss of correct conformation of the RING domain.

(ii) The protein equivalent of the del AG185 mutation which results in a frame shift at amino acid 22 followed by 17 out-of-frame amino acids and a stop codon. The BRCA1-delAG185 mutant was generated by PCR using the BRCA1-RF-5' oligonucleotide [SEQ ID NO: 14] and a 3' oligonucleotide that encoded the changed amino acid sequence: 5'-GCATGGATCCTCAAACCTTGTGCAGGCAGGTACCCTG
10 GTCAACAGGAGACAGGTGGGAAACCAGGATCTTTTGCATAGC-3' [SEQ ID NO: 29]. The truncated protein generated by the delAG185 mutation is found in high frequency in the Ashkenazi population [Struewing et al, Nat. Genet., 11:198-200
15 (1995)].

(iii) A truncated BRCA1 RING finger at amino acid 31, the result of a PCR error. The BRCA1-del31 truncation mutant was a mis-primed PCR reaction of BRCA1-RF identified by sequencing during the initial
20 screens for a wild-type LexA-BRCA1. The BRCA1-RF-trunc control is a truncation of the BRCA1-RF, a protein of 35 amino acids which ends within the first loop of the RING domain.

(iv) The RPT-1 RING finger domain. The
25 LexA-RPT-1 construct (amino acids 1-100) [SEQ ID NO: 5] was made by PCR-mediated amplification of the nucleotides representing the first 100 amino acids of the transcription factor RPT-1 [Patarca et al, Proc. Natl. Acad. Sci. USA, 85:2733-2737 (1988); RPT-1 cDNA kindly
30 provided by Dr. H. Cantor] with the 5' and 3' primers incorporating EcoRI and Sal I restriction sites.

(v) A non-specific control LexA fusion with RhoB. LexA-RhoB was a kind gift of Dr. George Prendergast, The Wistar Institute of Anatomy and Biology,
35 Philadelphia, PA.

All LexA mutant fusion constructs were made, as described for the wild-type BRCA1-RF, by cloning the appropriate mutated BRCA1-RF domain into the vector pBTM-116. The RPT-1 PCR product was enzymatically digested and ligated into the corresponding sites in pBTM-116. All clones were confirmed by sequencing.

The wild-type BRCA1-RF did not display intrinsic transcriptional activation function in yeast and proper expression of each LexA fusion in yeast was confirmed by Western blot analysis with anti-LexA DNA-binding domain antibody (data not shown).

These controls screen for proteins which interact only with wild-type the BRCA1-RF and not with any of the physiologically relevant BRCA1-RF mutations nor a RING finger that is the most similar to that of the BRCA1-RF [Miki et al, Science, 266:66-71 (1994)]. Thus, the controls make it possible to identify proteins which interact specifically with the BRCA1-RF and not with any other RING domain.

Example 2 - Yeast Two-Hybrid Screen for BRCA1-RF Interacting Proteins

To identify the potential protein partners of BRCA1, a yeast 2-hybrid analysis system as modified by Stan Hollenberg [Vojtek et al, cited above] was performed using the RING finger domain of human BRCA1. Guided by the expression patterns of BRCA1 during mouse development and in human spleen, the cDNA libraries selected for screening with the LexA-BRCA1-RF of Example 1 were (1) the human adult B cell, oligo-dT-primed, cDNA library [Durfee et al, Genes & Devel., 7:555-569 (1993) (a kind gift of Dr. Steve Elledge)] and (2) a whole mouse embryo (9.5-10.5 day), random-primed, cDNA library size selected for inserts of 300 to 500 base pairs in length [Vojtek et al, cited above; kind gift of Dr. Stan Hollenberg)].

Briefly, the LexA-BRCA1-RF and a selected library were co-transformed into the L40 yeast strain. Positive protein interactions were selected by His auxotrophy. Fifty colonies were picked and grown for 10 generations without selection for the LexA-BRCA1-RF plasmid. Isolated clones of each colony, which were positive for the presence of only the library plasmid, were picked and mated with AMR70 yeast containing LexA-BRCA1-RF, one of its mutants, or one of the LexA controls of Example 1A. Positive matings were selected by growth on media requiring the presence of both plasmids. These colonies were then scored for LacZ production (positive interaction) and those which were positive for interaction with the wild-type BRCA1-RF, but not any of the controls, were processed for further analysis.

Using the above assay methods, one hundred yeast colonies (50 from each library; each screen representing approximately $8-10 \times 10^6$ independent cDNAs), randomly taken from approximately 5-700 total colonies which grew on solid media lacking the amino acid histidine, were selected for additional screening.

Thirty-one cDNAs which specifically interacted with BRCA1-RF were obtained from the secondary screen of the two libraries. Eight of these (3 from the human library and 5 from the mouse library) encoded the same amino acid sequences.

A representative secondary screen of one of the human clones, hBAP-1 (aa483-729; SEQ ID NO: 6), and 3 of the mouse clones, mBAP-1 (aa581-720; SEQ ID NO: 8), mBAP-1 (aa518-(del)-718; SEQ ID NO: 7), and mBAP-1 (aa596-721; SEQ ID NO: 9) was performed by re-introducing the purified pACT plasmids containing them into naive yeast. The sequences of these clones are compared in FIG. 2B. This screen showed that each clone showed a strong interaction with the wild-type BRCA1 ring-finger,

but failed to interact with the C64G, C61G, del31, delAG, RPT-1, RhoB, or any of the specificity control LexA fusions (data not shown).

Thus, these clones specifically interact with only
5 the BRCA1 RING finger. These cDNA clones all encode the same region of the same protein which has been termed BRCA1-Associated Protein-1 or BAP-1. Each clone shares the same translational reading frame with the transcriptional activation domain to which it is fused.
10 In addition, the fusion junctions were different among the clones suggesting that the interaction was not due to a fusion-junction artifact.

The longest cDNA retrieved in the two-hybrid screen was a 2.0 kb clone from the human library and encoded 246
15 amino acids followed by a 1.3 kb 3'UTR. Each mouse clone encoded an overlapping, smaller subset of this human open reading frame and which served to partially map the minimal interaction domain. Further definition of this minimal interaction domain was performed by mutagenesis
20 of this region of BAP1.

The "minimal interaction domain" was determined by the shortest mouse clone [mBAP-1 (aa596-721; SEQ ID NO: 9)]. To further define the specificity of interaction between BRCA1 and BAP-1, carboxy- and amino-terminal
25 truncation mutants of mBAP-1 were generated by PCR-based deletion or point mutagenesis.

The appropriate region of mBAP1(596-721) was amplified by PCR using a vector primer
pVP16 5'-primer, 5'-CCGATGCCCTTGGAATTGACGAG-3';
30 pVP16 3'-primer, 5'-CGATGAATTCGAGCTAGCTTCTATC-3') and the appropriate truncating oligonucleotide
Mc43Ct1, 5'-GCATGAATTCTCAGCT CCGGCGCACTGAGATG-3';
Mc43Ct2, 5'-GCATGAATTCTCAAGCCAGCATGGATATGAAGG-3';
Mc43Ct3, 5'-GCATGAATTCTCAGTCATCAATCTTGAAGTTC-3';

Mc43Ct4, 5'-GCATGAATTCTCATGCAATCTCGGCTTCTAC-3'; or
Mc43Nt1, 5'-GCATG GATCCCCAAGATTGATGACCAGCGAAGG-3' [SEQ ID
NOS: 30 to 36, respectively].

These oligonucleotides were generated with an
5 incorporated EcoRI restriction site (for the 3' end
oligos) or a BamHI restriction site (for the 5' end
oligos). After PCR amplification, the product was cut
with BamHI and EcoRI, and then ligated into the mouse
library-yeast expression vector, pVP16 [Vojtek et al,
10 cited above].

The point mutant mBAP-1 (L691P) was made by standard
PCR-based mutagenesis protocols [Ho et al, Gene, 77:51-59
(1989)], using (Mc43(L691 P) sense-primer, 5'-
GCTGGCCAACCCGGTGGAAACAG-3' [SEQ ID NO: 37]; Mc43(L691P)
15 antisense-primer, 5'-CTGTTCCACCGGGTTGGCCAGC-3' [SEQ ID
NO: 38] and using the same vector primers described
above.

The "minimal interaction domain" was deleted from
the human sequence (the longest clone) and this protein
20 (hBAP-1(483-594) [SEQ ID NO: 6] was also assayed for
interaction with the BRCA1-RF in the yeast 2-hybrid
system.

All clones were confirmed by sequencing and
expression in yeast was confirmed by western analysis
25 using antibodies against the VP16 activation domain (data
not shown). Each individual mutant was co-transformed
with LexA-BRCA1-RF into L40 yeast and tested for
interaction via its ability to activate transcription
from the LacZ locus.

30 The mutants showed that deletion of protein sequence
from the carboxy or amino termini of mBAP-1 (aa 596-721;
SEQ ID NO: 9) almost completely destroyed the BAP1-BRCA1
interaction, suggesting a complex interface between the
proteins. Deletion of the last 20 amino acids of mBAP-1
35 led to a significant reduction in the intensity of

interaction. Further deletions from the COOH-terminus led to the complete loss of interaction between BRCA1-RF and BAP-1. A single amino-terminal truncation which deleted approximately half of mBAP-1 (aa 596-721; SEQ ID NO: 9) led to an almost complete loss of interaction. Interestingly, the mBAP(518del718) clone interacted most poorly with BRCA1-RF and lacked a 93 bp sequence (the reading frame was maintained), possibly the result of a naturally occurring splice variant. That these clones also fail to bind multiple, independent tumor-derived mutations of the BRCA1-RF provides strong genetic evidence for their relevance to the functions of BRCA1.

The results of the above experiments suggested that some critical domain was being disrupted by these truncations. A careful analysis showed that the region from amino acids 632 to 729 of SEQ ID NO: 6 may in fact generate a coiled-coil domain. A point mutation in the middle of the domain (leucine 691 substituted with a proline) destroys interaction with the BRCA1 RING structure. This result is consistent with the BAP-1/BRCA-1 interaction domain being a coiled-coil.

Example 3 - Analysis of BAP-1 cDNA

A nearly full-length cDNA was constructed via a combination of cDNA library screening, EST database searching, 5'RACE and RT-PCR (FIGs. 3A-3E) as follows. Searches of the protein and DNA databases [Altschul et al, J. Mol. Biol., 215:403-410 (1990)] with the BAP-1 protein/cDNA sequences obtained from the screening of Example 2, showed no significant matches with any known protein or cDNA. However, searches of the EST databases with BAP-1 cDNA yielded several "hits", including one whose clone had a 5' sequence that overlapped with the 3' sequence of another EST clone. The clones defined by these EST's were obtained from the I.M.A.G.E. consortium

[Lennon et al, Genomics, 33:151-152 (1996); clones #46154 and #40642]. A partial BAP-1 cDNA clone (EST-BAP1) was generated by digesting clone #40642 with Hind III and Fsp I and clone #46154 with Fsp I and EcoRI. These two
5 pieces were then ligated into the Hind III and EcoRI sites of the vector pcDNA3 (Invitrogen).

Analysis of the IMAGE consortium cDNA and its open-reading-frames suggested that this BAP-1 cDNA, as constructed, was not complete. Reverse-transcriptase-PCR
10 was performed on RNA from normal human fibroblasts using a gene-specific primer: 5'-GAAGCGGATGTCGTGGTAGG-3' [SEQ ID NO: 43] and identified 62 nucleotides which were missing from the "EST-BAP1" cDNA. These 62 nucleotides were inserted into the "EST-BAP1" cDNA by digestion of
15 the RT-PCR product with the restriction enzymes KpnI, which is a unique site within the 5' RT-PCR oligonucleotide:

5'CCTGTTATTAACCCTCACTAAAGGGAAGGGTACCATGAATAAGGGCTGGCT
GGAGC-3' [SEQ ID NO: 39] and 3' RT-PCR-oligonucleotide:
20 5'-GAAGCGGATGTCGTGGTAGG-3' [SEQ ID NO: 40], and AvrII. Ligation of the KpnI/AvrII digested RT-PCR fragment and AvrII/EcoRI digested "EST-BAP1" cDNA and the KpnI-EcoRI digested pcDNA3, produced the full-length BAP-1 cDNA.

Thus, BAP1 cDNA [FIGS. 3A to 3E; SEQ ID NO: 1]
25 comprises 3525 bp, including a polyA tract with multiple polyA signals. Conceptual translation yields a long open reading frame of 729 amino acids [SEQ ID NO: 2] with a predicted MW of about 81 kDa and pI of 6.3.

The presumptive initiator methionine is within a
30 favorable context for translation start, however the short 5'UTR of 39 bp encodes amino acids in-frame with the presumptive methionine and does not contain a stop codon. BLAST searches and a domain analysis [Henikoff & Henikoff, Genomics, 19:97-107 (1994)] indicated that BAP1
35 is a novel protein with motifs suggestive of function.

The amino-terminal 1-240 amino acids of SEQ ID NO: 2 show significant homology to a class of thiol proteases, designated ubiquitin C-terminal hydrolase (UCH), particularly Isozyme L3, which are implicated in the proteolytic processing of ubiquitin [Wilkinson et al, Science, 246:670-673 (1989)]. These enzymes play a key role in protein degradation via the ubiquitin-dependent proteasome pathway. The most closely related UCH is a hypothesized protein from *C. elegans* UCH-CAEEL, which shares 63% similarity (40% identity) with BAP1 through the UCH domain and is also likely to be a UCH enzyme. Pairwise similarities to other mammalian UCHs of 54% (UCHL3) and 56% (UCHL1) have also been found. Most importantly, the residues which form the catalytic site of BAP1 (Q85, C91, H169, and D184 of Figs. 3A-3E; SEQ ID NO: 2) are completely conserved, including the FELDG motif [Larsen et al, Biochemistry, 35:6735-6744 (1996)]. In addition, a loop of highly variable sequence, which is disordered in the crystallographic structure of human UCH-L3 [Johnston et al, EMBO J., 16:3787-3796 (1997)], is present (residues 140 to 167 of SEQ ID NO. 2). This loop may occlude the active site or provide substrate specificity for the enzyme.

BAP1 has a number of additional motifs; a region of extreme acidity spanning amino acids 396 to 408 of SEQ ID NO. 2, as well as multiple potential phosphorylation sites and N-linked glycosylation sites. The C-terminal one-third is highly charged and is rich in proline, serine and threonine. The extreme C-terminus contains two putative nuclear localization signals, KRKKFK and RRKRSR (aa 656-661 and aa 717-722 of SEQ ID NO: 2), and is hydrophilic; it is predicted to fold into a helical (possibly coiled-coil) structure. Indeed, within the BAP1 minimal interaction domain, (i.e., from about amino acid 596 to 729 of SEQ ID NO: 2) the mutation of leucine

691 to a proline, a change predicted to disrupt the helical nature of this region, abolished the BAP1-BRCA1 interaction, consistent with the hypothesis that BAP1 uses a coiled-coil domain to interact with the RING
5 finger domain of BRCA1. This overall architecture suggests that BAP1 is a new, structurally complex, and nuclear localized member of the UCH enzyme family.

Example 4 - In Vitro Protein Association

The direct interaction of the BRCA1-RF with BAP-1
10 was confirmed by binding of the BRCA1-RF to the fusion proteins of glutathione-S-transferase with BAP1.

A. The BAP/GST constructs

The original B cell library two-hybrid BAP-1 clone obtained from the screening experiments described
15 in Example 2 was pACT-hBAP1(483-729), which contained BAP1 amino acids 483-729 (nucleotides 1486 to 3525) [SEQ ID NOs: 2 and 1, respectively] in the pACT plasmid backbone. The glutathione S-transferase/BAP1 fusion protein, GST-hBAP1(483-729 of SEQ ID NO: 2), was
20 generated by cloning nucleotides 1486 to 3525 of SEQ ID NO: 1 from that original clone into pGEX-5x-1 (Pharmacia Biotech, Inc.).

Another BAP1 construct which lacked the minimal BRCA1 interaction domain pACT-hBAP1(483-594 of SEQ ID NO:
25 2), was generated and amplified by PCR using a pACT 5' vector primer 5'-GATGTATATAACTATCTATTTCG-3' [SEQ ID NO: 41] and the BAP 1-trunc. oligonucleotide: 5'-GCATAGATCTT CACCCCTGGCTGCCTTGGATTGG3' [SEQ ID NO: 42], which amplifies BAP1 nucleotides 1486-1821 of SEQ ID NO: 1.
30 The resulting sequence was digested with restriction enzymes and ligated into the vector pACT. Another fusion protein GST-hBAP1(483-594 of SEQ ID NO: 2) lacking the minimal BAP1 interaction domain, was generated in the

same manner as pACT-hBAP1(483-594 of SEQ ID NO: 2), described above, but fused to GST.

GST, and the BAP1 fusion constructs GST-hBAP1(483-729 of SEQ ID NO: 2) and GST-hBAP1(483-594 of SEQ ID NO: 2), were expressed in *E. coli* and then purified [Frangioni et al, Anal. Biochem., 210:179-187 (1993)]. ³⁵S-LexA-BRCA1-RF and ³⁵S-BRCA1 were produced in vitro via coupled transcription/translation (TNT®, Promega Corp., Madison, WI) in the presence of ³⁵S-Met.

10 B. Association Assay

Association between the proteins was assayed essentially as described by Barlev et al, J. Biol. Chem., 270:19337-19344 (1995). Briefly, each GST resin was incubated with the LexA-BRCA1-RF in 100 µL of incubation buffer (PBS containing 0.2 mM ZnSO₄, 0.05% NP-40 and 1 mM PMSF) for 1 hour at 4°C followed by a second hour at room temperature. The resin and associated proteins were then washed in incubation buffer twice (1 mL at room temperature for 15 minutes) followed by four washes in 20 PBS containing 300 mM NaCl, 0.2 mM ZnSO₄, 0.1% NP-40 and 1 mM PMSF. The associated proteins which remained bound to resin were eluted from the resin two times (15 minutes), each with 250 µL of elution buffer (100 mM TRIS, pH 8.0, 150 mM NaCl, 0.1% NP-40, 20 mM reduced glutathione). The two elutions were combined, concentrated to a volume of approximately 20 µL of a 50:50 resin slurry, and analyzed by SDS-PAGE and visualized by Coomassie blue staining and fluorography.

Association of the BRCA1-RF with BAP1 was confirmed in vitro by specific binding of ³⁵S-labeled LexA-BRCA1-RF to GST-hBAP1(483-729 of SEQ ID NO: 2) fusion protein, but not to GST alone, confirming a physical association of the two proteins.

To confirm that the association of the BRCA1-RF to BAP1 was not an artifact of using only a portion of

BRCA1, full length BRCA1 was expressed in vitro and incubated with GST and GST-hBAP1(483-729 of SEQ ID NO: 2). As a further control for the specificity of the interaction, BRCA1 was also incubated with
5 GST-hBAP1(483-594 of SEQ ID NO: 2), the GST-BAP1 fusion protein lacking the minimal interaction domain.

The BRCA1 protein specifically bound to GST-hBAP1(483-729 of SEQ ID NO: 2) and not to GST or GST-hBAP1(483-594 of SEQ ID NO: 2), confirming the direct
10 interaction of BRCA1 with BAP1 through the C-terminal region of BAP1.

Example 5 - Generation of Antibodies

Oligonucleotide primers (pACT 5'-vector primer 5'-GATGTATATACTATCTATTCG-3' [SEQ ID NO: 44]; BAP1 3'
15 primer (antibody) 5-CGTAGTCGACTGTCAGCGCCAGGGGACTC-3' [SEQ ID NO: 45]), were used to amplify the portion of the BAP1 cDNA [SEQ ID NO: 1] corresponding to amino acids 483 to 576 of SEQ ID NO: 2 via PCR cloning. The PCR product was then digested with the appropriate restriction enzymes
20 and ligated to the COOH-terminus of 6 Histidine residues of the vector pQE-30 (QIAGEN Inc.).

The His-tagged protein was purified from E. coli over a Ni-agarose column as described [Friedman et al, cited above] and was used to immunize rabbits for the
25 production of polyclonal antibodies (Cocalico Biologicals, Inc.).

Example 6 - Protein Expression of BAP1

COS-1 cells were grown at 37°C, 5% CO₂ in DMEM supplemented with 10% fetal bovine serum and 2mM
30 L-glutamine. COS1 cells were transiently transfected using DOSPOR transfection reagent (Boehringer Mannheim Biochemicals) following the manufacturers protocol with plasmids containing the BAP1 cDNA, e.g., pACT-hBAP1(483-

729 of SEQ ID NO: 2). The BAP1 cDNA was transcribed and translated in vitro in the presence of ³⁵S-Methionine. ³⁵S-labeled cytosolic and nuclear extracts were then prepared from transiently transfected COS1 cells.

5 Immunoprecipitation of BAP1 was performed by previously described procedures for the metabolic labeling and immunoprecipitation of proteins from cell lysates [Morris et al, Oncogene, 6:2339-2348, (1991); Rauscher et al, Science, 240:1010-1016 (1988); Friedman
10 et al, cited above] with either pre-immune or anti-BAP1 sera described in the above example.

As a control for nuclear localization, KAP-1, a co-repressor of transcription localized to the nucleus [Friedman et al, Genes Dev., 10:2067-2078, (1996)], was
15 also immunoprecipitated from these cell fractions. Immunoprecipitation of this product with anti-BAP-1 antiserum confirmed that the protein expressed in vitro from the cDNA resulted in a polypeptide that contained the antigen used to raise the antibodies produced as
20 described above. BAP-1 was found primarily in the nuclear fraction although a significant amount was detected in the cytosol. However, this may be an artifact of the cell fractionation procedure, since KAP-1 was also found to be present in both cytosolic and
25 nuclear fractions and in approximately the same ratio as BAP-1.

The expression of the BAP-1 cDNA in COS1 cells in vitro followed by immunoprecipitation of ³⁵S-labeled whole cell extract and analysis by SDS-PAGE also yielded
30 a single major protein with an apparent molecular weight of about 91 kDa. However, the largest BAP1 open reading frame encodes a protein of about 81 kDa predicted molecular weight. The difference between apparent and

predicted molecular weights may be accounted for by unusual properties of the C-terminus or by post-translational modifications.

Example 7 - Tissue and Cellular Expression of BAP-1

5 A. BAP1 is Expressed in a Variety of Tissues

 The direct interaction between BAP1 and BRCA1 illustrated in Example 4, suggests that BAP1 might be expressed in an overlapping subset of tissues expressing BRCA1 and that the subcellular location of BAP1 and BRCA1
10 may be the same.

 The expression of BAP1 in a variety of human adult tissues was determined by Northern blot analysis. Northern blot hybridizations were performed as follows: Ten μ g total RNA from multiple tissue RNA blots (Clontech
15 Laboratories, Inc., Palo Alto, CA), was electrophoretically gel-fractionated and transferred to Hybond N+ membranes (Amersham). The tissues represented were heart, brain, placenta, lung, liver, skeletal muscle, kidney, pancreas, spleen, thymus, prostate,
20 testis, ovary, small intestine, colon and peripheral blood lymphocytes.

 The protocols for hybridization of cDNA probes to RNA were performed as described (Clontech Laboratories, publication PR48380). Blots were
25 hybridized with a 2.0kbp 32 P-labeled hBAP1 cDNA (aa483-729; nucleotides 1486 to 3525) followed by washes under standard conditions and detection by autoradiography. Blots were also subsequently probed with a muscle actin cDNA.

30 The results indicated that the mRNA encoding BAP1 was present as a single mRNA species of about 4 kb in all tissues except testis, where a second, about 4.8 kb mRNA, was also detected. Highest expression was detected in testis, placenta and pancreas with varying

levels detected in the remaining tissues. Expression of BAP1 in normal breast tissue was confirmed by RT-PCR of total RNA isolated from normal human mammary epithelial cells (HUMEC; data not shown). The level and pattern of
5 tissue expression shown by BAP1 is similar to that shown by BRCA1 [Miki et al, cited above].

Northern blot analysis was also performed on several tumor cell lines representing a variety of tissue types. The cell line RNA blot was prepared by standard
10 methods (Sambrook et al, cited above) with 20 µg of total RNA. Equivalent loading of RNA was confirmed by ethidium bromide staining. Hybridization of cDNA probes to RNA were performed using the Clontech protocols. This hybridization also showed a single mRNA species. The
15 colon cell lines HT29 [ATCC HTB 28] and SK-Co-1 [ATCC HTB 39] showed no BAP-1 mRNA, suggesting some defect in the BAP-1 gene in these particular cell lines since colon tissue shows good expression of BAP-1.

B. BAP1 is a Nuclear Protein

20 The location of BAP1 as a nuclear protein within the cell was determined by immunofluorescence microscopy performed as previously described [Ishov et al, J. Cell Biology, 134:815-826 (1996)]. HEP2 epithelial cells were grown at 37°C, 5% CO₂ in DMEM
25 supplemented with 10% fetal bovine serum (FBS) and 2mM L-glutamine, and cells were transfected using DOSPOR transfection reagent (Boehringer Mannheim Biochemicals) following the manufacturers protocol via electroporation with the pcDNA3 vector (Invitrogen, Inc.) carrying the
30 BAP1 cDNA.

Transfectants were analyzed by immunofluorescence staining with anti-BAP1 polyclonal antibodies, which in turn, were detected with FITC using biotin-avidin enhancement. Cells were stained for DNA
35 with bis-benzimide (Hoechst 33258, Sigma Chemical Co.)

and mounted using Fluoromount G (Fisher Scientific). Analysis was performed with a confocal scanning microscope (Leica, Inc.).

Detection of BAP1 by confocal microscopy
5 located BAP1 almost exclusively in the nucleus of the cell consistent with its association with BRCA1, and the presence of two nuclear localization signals in the BAP1 protein sequence.

C. BAP1 is Located on Chromosome 3p21.3 and is
10 Mutated in Non-Small Cell Lung Carcinoma.

To determine whether BAP1 was located at a chromosomal region routinely mutated in breast cancer and thus may be a tumor suppressor gene, the deletion of which plays a critical role in tumor pathogenesis,
15 full-length BAP-1 cDNA was used in fluorescent in situ hybridization (FISH) of partial metaphases. FISH was performed as described previously [Tommerup and Vissing, Genomics, 27:259-264 (1995)] using a biotin-labelled 3.5 kb cDNA (full-length) clone of BAP-1, with corresponding
20 DAPI-stained chromosome banding. Localization of BAP1 was based on the DAPI-band pattern and measurement of the relative distance from the short arm telomere to the signals (FLpter value).

BAP1 maps to chromosome 3p21.3. Specific
25 signals were observed only on the midportion of the short arm of chromosome 3 with 42 of 69 analyzed metaphase spreads showing at least one specific signal. The FLpter value was 0.27 ± 0.02 , corresponding to a localization for BAP1 at 3p21.2-p21.31. This location is a region of
30 LOH for breast cancer as well as a region frequently deleted in lung carcinomas [Buchhagen et al, Int. J. Cancer, 57:473-479 (1994); Thiberville et al, Int. J. Cancer, 64:371-377 (1995)].

Example 8 - Mutational Analysis of BAP1

The chromosomal location of BAP1 suggested the possibility of mutations within BAP1 in lung and breast tumors. Thus, a variety of tumor cell lines were
5 screened for mutations within the BAP1 gene by Southern, Northern and PCR-based SSCP analyses.

A. RNA/DNA Preparation

Genomic DNA from a panel of small cell lung cancer (SCLC), non-small cell lung cancer (NSCLC), breast
10 cancer, and lymphoblastoid cell lines was prepared using standard methods. All cell lines were identified by their NCI number [Phelps et al, J. Cell. Biochem. Suppl., 24:32-91 (1996)]: H727, H1466, H226, H526, H841, H1045, H289, BL1672, BL1770, H289, H847, H920, H1450, H1573,
15 H1155, H1299, H1693. Total RNA was extracted by the cesium chloride ultracentrifugation method [Ausubel et al, Current Protocols in Molecular Biology, J. Kaaren ed., John Wiley & Sons, Inc. (1987)]. First strand cDNAs were synthesized from RNA by M-MLV reverse transcriptase
20 (Gibco BRL) according to the manufacturer's instructions.

B. Single Strand Conformational Polymorphism (SSCP) Analysis

Seventeen overlapping PCR primer pairs, each with a predicted product size of approximately 200 base
25 pairs, were designed to span the 2.2 kb open reading frame of the BAP1 cDNA sequence. cDNA (from RNA) was amplified in 20 µl PCR reactions containing 20mM Tris HCl (pH 8.3), 50mM KCl, 1.5 mM MgCl, 0.2 mM each dNTP, 0.1 mM each forward and reverse primer, 0.05 ml 32P-α dCTP, and
30 0.5 units Taq DNA Polymerase (BRL). PCR reactions were carried out in a Perkin-Elmer 9600 Thermocycler using a touchdown technique: a 2.5 minute initial denaturation at 94°C was followed by 35 cycles of denaturation at 94°C x 30s, annealing, initially at 65°C decreasing by 1°C for
35 each of the first ten cycles to 55°C, x 30s, and

extension at 72°C x 30s with a final extension of 5 minutes at 72°C. PCR products were then diluted 1:10 with SSCP dye (95% formamide, 20mM EDTA, and 0.05% each of bromophenol blue and xylene cyanol), heat denatured, and electrophoresed on 0.5X MDE gels +/- 10% glycerol. Abnormal single stranded DNA detected as autoradiographic shifts were reamplified by PCR and subjected to automated dye-terminator sequencing (ABI 373).

SSCP analysis showed a homozygous shift in H1466 detected by RT-PCR amplification spanning nts 1089 to 1286 (primers: sense 5'-CAACCCCACTCCCATTGTC-3" [SEQ ID NO: 46]; antisense 5'-GAGTTGGTGTCTGCACGTC-3" [SEQ ID NO: 47]). Automated sequencing revealed a homozygous 8 base pair frameshift deletion in the NCI-H1466 cDNA, predicted to encode a truncated 393 amino acid BAP1 protein. This homozygous deletion was confirmed to be present in genomic DNA from the same cell line. In the NCI-H226 line, only the 2.4 kb band and an aberrant 2.6 kb band were detected.

B. Northern Analysis

These cell lines were subjected to Northern blot analysis and EcoRI digestion and then hybridized to a full-length BAP1 cDNA probe. A single 23 kb band was detected in the lymphoblastoid and most tumor cell lines (data not shown). One NSCLC line, NCI-H226, did not show the 23 kb band but did show an aberrant 30 kb band (data not shown).

Further mutational analysis was performed by screening a panel of lung cancer and lymphoblastoid cell lines for expression of BAP1 mRNA. Northern blot hybridization showed that most cell lines expressed a single 4 kb mRNA. A fainter (5.0 kb) band was visible corresponding to cross-hybridization with the 28S ribosomal component. However, two cell lines, NCI-H226 and the non-small cell lung cancer NCI-H1466 (both

NSCLCs), showed undetectable levels of BAP1 expression, suggesting that BAP1 may play a critical role in NSCLC pathogenesis.

C. Southern Analysis

5 To further characterize this potential genomic rearrangement, genomic DNA from NCI-H226 and a smaller number of lung cancer and lymphoblastoid lines were subjected to Southern blot hybridization. Briefly, five
10 μg of genomic DNA was subjected to restriction enzyme digestion with BamHI. Using the full-length BAP1 cDNA probe, four distinct bands at 7.5 kb, 4.0 kb, 3.0 kb, and 2.4 kb were detected which were present in all cell lines tested with the exception of NCI-H226. The non-small
15 cell lung cancer NCI-H226 line shows an absence of the 7.5kb, 4.0kb, and 3.0kb bands. An aberrant 2.6kb band is detected in the NCI-H226 cell line.

 These data clearly show that genetic alterations, including intragenic homozygous deletions, occur in BAP1.

EXAMPLE 9: BAP1 Augments the Growth Suppressive Activity of BRCA1

20 To determine whether BAP1 may affect cell growth itself or may affect BRCA1-mediated changes in cell growth, BRCA1 and BAP1 cDNAs were co-transfected into MCF7 breast cancer cells. This cell line was chosen for
25 several reasons. It has been previously shown that these cells are inhibited by the overexpression of BRCA1 [Holt et al, cited above]. Both northern and RT/PCR analyses showed that BAP1 was expressed in this cell line (data not shown); and analysis of the open reading frame from
30 BAP1 cDNA prepared from this cell line showed no mutations (data not shown).

 MCF7 cells grown at 37°C, 5% CO2 in DMEM supplemented with 10% FBS and non-essential amino acids, were transfected with the following plasmid pairs:

(a) empty plasmids pcDNA3 and pCMV5; (b) pcDNA3 and pCMV5-BAP1; (c) pcDNA3 and pCMV5-BAP1(165-729 of SEQ ID NO: 2); (d) pcDNA3-BRCA1 and pCMV5; (e) pcDNA3-BRCA1 and pCMV5-BAP1; (f) pcDNA3-BRCA1 and pCMV5-BAP1(165-729 of
5 SEQ ID NO: 2); (g) pcDNA3-BRCA1- Δ 11 and pCMV5; (h) pcDNA3-BRCA1- Δ 11 and pCMV5-BAP1; and (i) pcDNA3-BRCA1- Δ 11 and pCMV5-BAP1(165-729 of SEQ ID NO: 2) by a modified CaPO_4 -DNA precipitation method [Holt et al, cited above].
MCF7 cells, at 2×10^6 cells/10 cm dish, were fed
10 fresh medium approximately 3 hours prior to transfection and were then treated with the Ca-DNA precipitate for 4 hours. The cells were subjected to a brief shock with transfection buffer containing 15% glycerol. Twelve to sixteen hours later, the cells were trypsinized, counted
15 and plated directly into complete medium containing 0.75 mg/mL G418 at 5×10^5 cells per 10 cm dish. Cells were fed fresh medium containing G418 every three to four days. Cells were stained for colonies approximately 21 to 28 days after transfection. The experiment was repeated 4
20 times with similar results.

The expression of BRCA1 alone (pcDNA3-BRCA1:pCMV5) decreased the number of colonies formed by these cells when compared to the empty vector control (pcDNA3:pCMV5), in agreement with other studies [Holt et al, cited
25 above]. The co-expression of BRCA1 and BAP1 (pcDNA3-BRCA1:pCMV5-BAP1) significantly decreased the number of cell colonies (approximately 4 fold vs. BRCA1 alone) indicating that BAP1 enhances the growth suppressive actions of BRCA1. A mutant of BAP1, BAP1(AA165-729), in
30 which the enzymatic region is deleted but which still binds to BRCA1 (data not shown), also enhanced the growth suppression of BRCA1, but not to the same extent as the wildtype BAP1.

In contrast to BRCA1, the expression of BRCA1- Δ 11
35 (BRCA1 missing the 11th exon) in MCF7 cells by itself had

no effect on the growth of MCF7 cells. However, the co-expression of BRCA1- Δ 11 and BAP1 significantly decreased the number of colonies, suggesting that the presence of BAP1 could functionally substitute for the missing 11th exon of BRCA1 and/or that BAP1 itself was an inhibitor of cell growth.

In support of this latter hypothesis, the expression of BAP1 in MCF7 cells did somewhat reduce the number of colonies formed (pcDNA3:pcMV5-BAP1). The expression of the enzymatic mutant, BAP1(165-729), alone or in combination with BRCA1- Δ 11 yielded the same number of colonies. Thus, enzymatically active BAP1 enhances BRCA1-mediated suppression of growth.

Example 10 - BAP1 Enzymatic Assay

To determine whether BAP1 did indeed have UCH activity, the BAP1 cDNA was expressed in bacteria and this protein was assayed for the ability to hydrolyze the glycine 76 ethyl ester of ubiquitin [Ub-OEt; Mayer et al, Biochemistry, 28:166-172 (1989)].

Briefly, bacteria (E. coli DH5a) harboring an IPTG-inducible expression plasmid containing BAP1 or an enzymatically null mutant, BAP1 (C91 S) (pQE-30; QIAGEN Inc.) were grown and induced with 1 mM IPTG for 4 hours. The bacteria were collected and the pellets were resuspended to 1/20 volume (original culture) in lysate buffer (50mM Tris, pH 8.0, 25mM EDTA, 10mM 2-mercapto-ethanol, 100 μ g/ml lysozyme). The lysates were sonicated and centrifuged at 40,000 Xg.

The pellets were resuspended in an volume equal to that of the supernatant and samples of both pellet and supernatant were analyzed by SDS-PAGE for expression levels and inclusion body formation. Induction of protein was verified by SDS-PAGE of each fraction.

Overexpression of BAP1 in bacteria led to abundant protein, most of which was found in an inactive, insoluble form.

Assays for BAP1 enzymatic activity, specifically, ubiquitin carboxy-terminal hydrolase activity, were performed on the above-described soluble fraction essentially as described for the UCH-L1 and UCH-L3 enzymes using the glycine 76 ethyl ester of ubiquitin (Ub-OEt) as a substrate [Mayer et al, cited above; Wilkinson et al, Biochemistry, 25:6644-6649 (1986)]. Assays were done in triplicate. The peak areas were integrated and normalized with respect to a ubiquitin standard.

The BAP1 protein found in the soluble fraction was able to hydrolyze UbOEt and the level of this activity increased with the level of protein, indicating that BAP1 contains UCH-like enzymatic activity.

The active site thiol residue responsible for UCH activity in UCH-L3 has been identified and its mutation leads to abolition of enzyme activity [Larsen et al, cited above]. Mutation of the corresponding cysteine residue in BAP1, BAP1 (C91 S), yielded a protein with no UCH activity, further suggesting that BAP1 is a thiol protease of the UCH family.

BAP1's identity as a protease of the ubiquitin carboxy-terminal hydrolase (UCH) family implies a role for either ubiquitin-mediated, proteasome dependent degradation or other ubiquitin-mediated regulatory [Isaksson et al, Biochimica et Biophysica Acta, 1288:F21-29 (1996)] pathways in BRCA1 function. Regulated ubiquitination of proteins and subsequent proteasome-dependent proteolysis plays a role in almost every cellular growth, differentiation and homeostatic process [reviewed by Ciechanover, Biol. Chem. Hoppe-Seyler, 375:565-581 (1994); Isaksson et al, cited above;

Wilkinson, Annual Review of Nutrition, 15:161-189

(1995)]. This pathway can be broadly subdivided into reactions involving 1) pro-ubiquitin processing and ATP-dependent activation of ubiquitin; 2) substrate
5 recognition, conjugation and editing of the polyubiquitin chain; 3) proteasome-dependent degradation of the ubiquitin protein and; 4) cleavage and/or debranching of peptide-ubiquitin conjugates and recycling of ubiquitin
10 to cellular pools. The pathway is regulated at almost every step. First, at the level of substrate specificity via the concerted actions of activating enzymes, carrier proteins and ligation enzymes, and secondly, at the level of proteolytic deubiquitination and ubiquitin hydrolysis.

The UCH family has been characterized as a set of
15 small (25-30 kDa) cytoplasmic proteins which prefer to cleave ubiquitin from ubiquitin-conjugated small substrates and may also be involved in the co-translational processing of proubiquitin. UCHs show considerable tissue specificity and developmentally-timed
20 regulation [Wilkinson et al, Biochem. Soc. Trans., 20:631-637 (1992)]. UCH family members are strongly and differentially expressed in neuronal, hematopoietic and germ cells in many species. Most remarkably, a novel UCH enzyme has recently been cloned from *Aplysia californica*
25 whose enzymatic function is essential for acquisition and maintenance of long-term memory [Hedge et al, Cell, 89:114-126 (1997)]. Finally, UCH levels are strongly downregulated during viral transformation of fibroblasts [Honore et al, FEBS Letter, 280:235-240 (1991)],
30 consistent with a role in growth control.

BAP1 is the newest member of the UCH family and considerably expands the potential roles of this family of proteases. BAP1 is a much larger protein (90 kDa) and is the first nuclear-localized UCH. BAP1 is also likely
35 to be involved in the regulation of protein subcellular

localization. Ubiquitin, or a ubiquitin-like moiety, may affect the specific targeting of proteins to locations other than the proteasome [see, e.g., Mahajan et al, Cell, 88:97-107 (1997)]. BAP1-mediated removal of
5 "ubiquitin" from BRCA1, or a protein associated with BRCA1, could target it for removal to another cellular compartment, thus functionally destroying the protein without physically doing so.

BRCA1 is also localized in nuclear dot structures in
10 a cell-cycle dependent manner [Scully et al, cited above]. This association of BRCA1 with RAD51 in both mitotic and meiotic cells broadly implicates BRCA1 in DNA repair and/or recombination processes. The
RAD51/52-dependent DNA repair pathway is highly regulated
15 and includes many proteins, some of which may be potential substrates for BAP1-mediated ubiquitin hydrolysis [Watkins et al, Molecular & Cellular Biology, 13:7757-7765 (1993)]. Thus, it appears that the DNA
repair machinery contains both ubiquitin-conjugating and
20 -hydrolyzing elements, since BAP1 is now implicated as a member of the BRCA1/RAD51/hUBC9 complex. It is possible that BAP1, which is co-expressed with BRCA1 in testis, may regulate the recombination/repair functions of the
BRCA1/RAD52 complex by targeting either RAD23 or UBL1 for
25 ubiquitin hydrolysis.

Numerous modifications and variations of the present invention are included in the above-identified
specification and are expected to be obvious to one of
skill in the art. Such modifications and alterations to
30 the compositions and processes of the present invention are believed to be encompassed in the scope of the claims appended hereto.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: Wistar Institute of Anatomy & Biology
Rauscher III, Frank J.
Jensen, David E.
- (ii) TITLE OF INVENTION: BRCA1 Associated Protein (BAP-1) and
Uses Therefor
- (iii) NUMBER OF SEQUENCES: 47
- (iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Howson and Howson
(B) STREET: P.O. Box 457, 321 Norristown Road
(C) CITY: Spring House
(D) STATE: Pennsylvania
(E) COUNTRY: U.S.A.
(F) ZIP: 19477
- (v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30
- (vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: WO
(B) FILING DATE:
(C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 60/022,997
(B) FILING DATE: 02-AUG-1996
- (viii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 60/038,109
(B) FILING DATE: 19-FEB-1997
- (viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Bak, Mary E.
(B) REGISTRATION NUMBER: 31,215
(C) REFERENCE/DOCKET NUMBER: WST68BPCT
- (ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 215-540-9200
(B) TELEFAX: 215-540-5818

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3517 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: cDNA

53

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION: 40..2226

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

GGCACGAGGC	ATGGCGCTGA	GGGGCCGCC	CGCGGGAAG	ATG Met 1	AAT Asn	AAG Lys	GGC Gly	TGG Trp 5	54
CTG Leu	GAG Glu	CTG Leu	GAG Glu	AGC Ser 10	GAC Asp	CCA Pro	GGC Gly	CTC Leu	102
TTC Phe	GGT Gly	GTC Val	AAG Lys 25	GGG Gly	GTG Val	CAA Gln	GTG Val	GAG Glu 30	150
AAA Lys	TGT Cys	CAG Gln 40	GGC Gly	CCT Pro	GTA Val	TAT Tyr	GGA Gly 45	TTT Phe	198
GAA Glu	GAG Glu	CGC Arg	CGG Arg	TCC Ser	CGG Arg	CGA Arg 60	AAG Lys	GTC Val	246
TCC Ser 70	GTG Val	ATT Ile	GAT Asp	GAT Asp 75	GAT Asp	ATT Ile	GTG Val	AAT Asn	294
CTG Leu	ATA Ile	CCC Pro	AAC Asn	TCT Ser 90	TGT Cys	GCA Ala	ACT Thr	CAT His	342
AAC Asn	TGC Cys	AGC Ser	AGC Ser 105	GTG Val	GAC Asp	CTG Leu	GGA Gly	CCC Pro 110	390
TTC Phe	ACC Thr	AAG Lys 120	GGT Gly	TTC Phe	AGC Ser	CCT Pro	GAG Glu 125	AGC Ser	438
GCC Ala 135	CCG Pro	GAG Glu	TTG Leu	GCC Ala	AAG Lys	GCC Ala 140	CAT His	AAT Asn	486
CGC Arg 150	CAC His	CTC Leu	CCT Pro	GAG Glu	AAG Lys 155	CAG Gln	AAT Asn	GGC Gly	534
GAG Glu	CGG Ala	TTC Phe	CAC His	TTT Phe 170	GTC Val	AGC Ser	TAT Tyr	GTG Val	582
GAG Glu	CTG Leu	GAT Asp	GGG Gly 185	CTG Leu	AAG Lys	GTC Val	TAC Tyr	CCC Pro 190	630
GAG Glu	GAC Asp	GAG Glu 200	GAG Glu	TGG Trp	ACA Thr	GAC Asp	AAG Lys 205	GCC Ala	678

54

ATC Ile	GGC Gly 215	CTC Leu	GCC Ala	ACT Thr	GCA Ala	GGG Gly 220	GAG Glu	CCC Pro	TAC Tyr	CAC His	GAC Asp 225	ATC Ile	CGC Arg	TTC Phe	AAC Asn	726
CTG Leu 230	ATG Met	GCA Ala	GTG Val	GTG Val	CCC Pro 235	GAC Asp	CGC Arg	AGG Arg	ATC Ile	AAG Lys 240	TAT Tyr	GAG Glu	GCC Ala	AGG Arg	CTG Leu 245	774
CAT His	GTG Val	CTG Leu	AAG Lys	GTG Val 250	AAC Asn	CGT Arg	CAG Gln	ACA Thr	GTA Val 255	CTA Leu	GAG Glu	GCT Ala	CTG Leu	CAG Gln 260	CAG Gln	822
CTG Leu	ATA Ile	AGA Arg	GTA Val 265	ACA Thr	CAG Gln	CCA Pro	GAG Glu	CTG Leu 270	ATT Ile	CAG Gln	ACC Thr	CAC His	AAG Lys 275	TCT Ser	CAA Gln	870
GAG Glu	TCA Ser	CAG Gln 280	CTG Leu	CCT Pro	GAG Glu	GAG Glu	TCC Ser 285	AAG Lys	TCA Ser	GCC Ala	AGC Ser	AAC Asn 290	AAG Lys	TCC Ser	CCG Pro	918
CTG Leu	GTG Val 295	CTG Leu	GAA Glu	GCA Ala	AAC Asn	AGG Arg 300	GCC Ala	CCT Pro	GCA Ala	GCC Ala	TCT Ser 305	GAG Glu	GGC Gly	AAC Asn	CAC His	966
ACA Thr 310	GAT Asp	GGT Gly	GCA Ala	GAG Glu	GAG Glu 315	GCG Ala	GCT Ala	GGT Gly	TCA Ser	TGC Cys 320	GCA Ala	CAA Gln	GCC Ala	CCA Pro	TCC Ser 325	1014
CAC His	AGC Ser	CCT Pro	CCC Pro	AAC Asn 330	AAA Lys	CCC Pro	AAG Lys	CTA Leu	GTG Val 335	GTG Val	AAG Lys	CCT Pro	CCA Pro	GGC Gly 340	AGC Ser	1062
AGC Ser	CTC Leu	AAT Asn	GGG Gly 345	GTT Val	CAC His	CCC Pro	AAC Asn	CCC Pro 350	ACT Thr	CCC Pro	ATT Ile	GTC Val 355	CAG Gln	CGG Arg	CTG Leu	1110
CCG Pro	GCC Ala	TTT Phe 360	CTA Leu	GAC Asp	AAT Asn	CAC His	AAT Asn 365	TAT Tyr	GCC Ala	AAG Lys	TCC Ser	CCC Pro 370	ATG Met	CAG Gln	GAG Glu	1158
GAA Glu	GAA Glu 375	GAC Asp	CTG Leu	GCG Ala	GCA Ala	GGT Gly 380	GTG Val	GGC Gly	CGC Arg	AGC Ser	CGA Arg 385	GTT Val	CCA Pro	GTC Val	CGC Arg	1206
CCA Pro 390	CCC Pro	CAG Gln	CAG Gln	TAC Tyr	TCA Ser 395	GAT Asp	GAT Asp	GAG Glu	GAT Asp	GAC Asp 400	TAT Tyr	GAG Glu	GAT Asp	GAC Asp	GAG Glu 405	1254
GAG Glu	GAT Asp	GAC Asp	GTG Val	CAG Gln 410	AAC Asn	ACC Thr	AAC Asn	TCT Ser	GCC Ala 415	CTT Leu	AGG Arg	TAT Tyr	AAG Lys	GGG Gly 420	AAG Lys	1302
GGA Gly	ACA Thr	GGG Gly	AAG Lys 425	CCA Pro	GGG Gly	GCA Ala	TTG Leu	AGC Ser 430	GGT Gly	TCT Ser	GCT Ala	GAT Asp	GGG Gly 435	CAA Gln	CTG Leu	1350
TCA Ser	GTG Val	CTG Leu 440	CAG Gln	CCC Pro	AAC Asn	ACC Thr	ATC Ile 445	AAC Asn	GTC Val	TTG Leu	GCT Ala	GAG Glu 450	AAG Lys	CTC Leu	AAA Lys	1398
GAG Glu	TCC Ser 455	CAG Gln	AAG Lys	GAC Asp	CTC Leu	TCA Ser 460	ATT Ile	CCT Pro	CTG Leu	TCC Ser	ATC Ile 465	AAG Lys	ACT Thr	AGC Ser	AGC Ser	1446

55

GGG Gly 470	GCT Ala	GGG Gly	AGT Ser	CCG Pro	GCT Ala 475	GTG Val	GCA Ala	GTG Val	CCC Pro	ACA Thr 480	CAC His	TCG Ser	CAG Gln	CCC Pro	TCA Ser 485	1494
CCC Pro	ACC Thr	CCC Pro	AGC Ser	AAT Asn 490	GAG Glu	AGT Ser	ACA Thr	GAC Asp	ACG Thr 495	GCC Ala	TCT Ser	GAG Glu	ATC Ile	GGC Gly 500	AGT Ser	1542
GCT Ala	TTC Phe	AAC Asn	TCG Ser 505	CCA Pro	CTG Leu	CGC Arg	TCG Ser	CCT Pro 510	ATC Ile	CGC Arg	TCA Ser	GCC Ala	AAC Asn 515	CCG Pro	ACG Thr	1590
CGG Arg	CCC Pro	TCC Ser 520	AGC Ser	CCT Pro	GTG Val	ACC Thr	TCC Ser 525	CAC His	ATC Ile	TCC Ser	AAG Lys	GTG Val 530	CTT Leu	TTT Phe	GGA Gly	1638
GAG Glu	GAT Asp 535	GAC Asp	AGC Ser	CTG Leu	CTG Leu	CGT Arg 540	GTG Val	GAC Asp	TGC Cys	ATA Ile	CGC Arg 545	TAC Tyr	AAC Asn	CGT Arg	GCT Ala	1686
GTG Val 550	CGT Arg	GAT Asp	CTG Leu	GGT Gly	CCT Pro 555	GTG Val	ATC Ile	AGC Ser	ACA Thr	GGC Gly 560	CTG Leu	CTG Leu	CAC His	CTG Leu	GCT Ala 565	1734
GAG Glu	GAT Asp	GGG Gly	GTG Val	CTG Leu 570	AGT Ser	CCC Pro	CTG Leu	GGC Ala	CTG Leu 575	ACA Thr	GAG Glu	GGT Gly	GGG Gly	AAG Lys 580	GGT Gly	1782
TCC Ser	TCG Ser	CCC Pro	TCC Ser 585	ATC Ile	AGA Arg	CCA Pro	ATC Ile	CAA Gln 590	GGC Gly	AGC Ser	CAG Gln	GGG Gly	TCC Ser 595	AGC Ser	AGC Ser	1830
CCA Pro	GTG Val	GAG Glu 600	AAG Lys	GAG Glu	GTG Val	GTG Val	GAA Glu 605	GCC Ala	ACG Thr	GAC Asp	AGC Ser	AGA Arg 610	GAG Glu	AAG Lys	ACG Thr	1878
GGG Gly	ATG Met 615	GTG Val	AGG Arg	CCT Pro	GGC Gly	GAG Glu 620	CCC Pro	TTG Leu	AGT Ser	GGG Gly	GAG Glu 625	AAA Lys	TAC Tyr	TCA Ser	CCC Pro	1926
AAG Lys 630	GAG Glu	CTG Leu	CTG Leu	GCA Ala	CTG Leu 635	CTG Leu	AAG Lys	TGT Cys	GTG Val	GAG Glu 640	GCT Ala	GAG Glu	ATT Ile	GCA Ala	AAC Asn 645	1974
TAT Tyr	GAG Glu	GCG Ala	TGC Cys	CTC Leu 650	AAG Lys	GAG Glu	GAG Glu	GTA Val	GAG Glu 655	AAG Lys	AGG Arg	AAG Lys	AAG Lys	TTC Phe 660	AAG Lys	2022
ATT Ile	GAT Asp	GAC Asp	CAG Gln 665	AGA Arg	AGG Arg	ACC Thr	CAC His	AAC Asn 670	TAC Tyr	GAT Asp	GAG Glu	TTC Phe	ATC Ile 675	TGC Cys	ACC Thr	2070
TTT Phe	ATC Ile	TCC Ser 680	ATG Met	CTG Leu	GCT Ala	CAG Gln	GAA Glu 685	GGC Gly	ATG Met	CTG Leu	GCC Ala	AAC Asn 690	CTA Leu	GTG Val	GAG Glu	2118
CAG Gln	AAC Asn 695	ATC Ile	TCC Ser	GTG Val	CGG Arg	CGG Arg 700	CGC Arg	CAA Gln	GGG Gly	GTG Val	AGC Ser 705	ATC Ile	GGC Gly	CGG Arg	CTC Leu	2166
CAC His 710	AAG Lys	CAG Gln	CGG Arg	AAG Lys	CCT Pro 715	GAC Asp	CGG Arg	CGG Arg	AAA Lys	CGC Arg 720	TCT Ser	CGC Arg	CCC Pro	TAC Tyr	AAG Lys 725	2214

56

GCC	AAG	CGC	CAG	TGAGGACTGC	TGGCCCTGAC	TCTGCAGCCC	ACTCTTGCCG	2266
Ala	Lys	Arg	Gln					
TGTGGCCCTC		ACCAGGGTCC	TTCCCTGCCC	CACTTCCCT	TTTCCCAGTA	TTACTGAATA		2326
GTCCAGCTG		GAGAGTCCAG	GCCCTGGGAA	TGGGAGGAAC	CAGGCCACAT	TCCTTCCATC		2386
GTGCCCTGAG		GCCTGACACG	GCAGATCAGC	CCCATAGTGC	TCAGGAGGCA	GCATCTGGAG		2446
TTGGGGCACA		GCGAGGTACT	GCAGCTTCCT	CCACAGCCGG	CTGTGGAGCA	GCAGGACCTG		2506
GCCCTTCTGC		CTGGGCAGCA	GAATATATAT	TTTACCTATC	AGAGACATCT	ATTTTCTGG		2566
GCTCCAACCC		AACATGCCAC	CATGTTGACA	TAAGTTCCTA	CCTGACTATG	CTTCTCTCC		2626
TAGGAGCTGT		CCTGGTGGGC	CCAGGTCTT	GTATCATCCA	CGGTCCCAAC	TACAGGGTCC		2686
TAGCTGGGGG		CCTGGGTGGG	CCCTGGGCTC	TGGGCCCTGC	TGCTCTAGCC	CCAGCCACCA		2746
GCCTGTCCCT		GTTGTAAGGA	AGCCAGGTCT	TCTCTCTTCA	TTCTCTTAG	GAGAGTGCCA		2806
AACTCAGGGA		CCCAGCACTG	GGCTGGGTG	GGAGTAGGGT	GTCCCACTGG	GGTTGGGGTG		2866
AGCAGGCTGC		TGGGATCCCA	TGGCCTGAGC	AGAGCATGTG	GGAAGTGTTC	AGTGGCCTGT		2926
GAACTGTCTT		CCTTGTCTA	GCCAGGCTGT	TCAAGACTGC	TCTCCATAGC	AAGGTCTAG		2986
GGCTCTTCGC		CTTCAGTGTT	GTGGCCCTAG	CTATGGGCCT	AAATTGGGCT	CTAGGCTCT		3046
GTCCCTGGCG		CTTGAGGCTC	AGAAGAGCCT	CTGTCCAGCC	CCTCAGTATT	ACCATGTCTC		3106
CCTCTCAGGG		GTAGCAGAGA	CAGGGTTGCT	TATAGGAAGC	TGGCACCACT	CAGCTCTTCC		3166
TGCTACTCCA		GTTTCCTCAG	CCTCTGCAAG	GCACTCAGGG	TGGGGGACAG	CAGGATCAAG		3226
ACAACCCGTT		GGAGCCCTG	TGTTCCAGAG	GACCTGATGC	CAAGGGGTAA	TGGGCCCAGC		3286
AGTGCCTCTG		GAGCCCAGGC	CCCAACACAG	CCCCATGGCC	TCTCCAGATG	GCTTTGAAAA		3346
GGTGATCCAA		CAGGCCCTT	TATCTGTACA	TAGTGACTGA	GTGGGGGGTG	CTGGCAAGTG		3406
TGGCACTCCT		CTGGGCTGAG	CACAGCTTGA	CCCTCTAGC	CCCTGTAAAA	CTGGATCAAT		3466
GAATGAATAA		AACTCTCCTA	AGATCTCCTG	AGAAAAAAGA	AAAAAAAAG	G		3517

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 729 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met	Asn	Lys	Gly	Trp	Leu	Glu	Leu	Glu	Ser	Asp	Pro	Gly	Leu	Phe	Thr
1				5					10					15	
Leu	Leu	Val	Glu	Asp	Phe	Gly	Val	Lys	Gly	Val	Gln	Val	Glu	Glu	Ile
			20					25					30		

57

Tyr	Asp	Leu 35	Gln	Ser	Lys	Cys	Gln 40	Gly	Pro	Val	Tyr	Gly 45	Phe	Ile	Phe
Leu	Phe 50	Lys	Trp	Ile	Glu	Glu 55	Arg	Arg	Ser	Arg	Arg 60	Lys	Val	Ser	Thr
Leu 65	Val	Asp	Asp	Thr	Ser 70	Val	Ile	Asp	Asp	Asp 75	Ile	Val	Asn	Asn	Met 80
Phe	Phe	Ala	His	Gln 85	Leu	Ile	Pro	Asn	Ser 90	Cys	Ala	Thr	His	Ala 95	Leu
Leu	Ser	Val	Leu 100	Leu	Asn	Cys	Ser	Ser 105	Val	Asp	Leu	Gly	Pro 110	Thr	Leu
Ser	Arg	Met 115	Lys	Asp	Phe	Thr	Lys 120	Gly	Phe	Ser	Pro	Glu 125	Ser	Lys	Gly
Tyr	Ala 130	Ile	Gly	Asn	Ala	Pro 135	Glu	Leu	Ala	Lys	Ala 140	His	Asn	Ser	His
Ala 145	Arg	Pro	Glu	Pro	Arg 150	His	Leu	Pro	Glu	Lys 155	Gln	Asn	Gly	Leu	Ser 160
Ala	Val	Arg	Thr	Met 165	Glu	Ala	Phe	His	Phe 170	Val	Ser	Tyr	Val	Pro 175	Ile
Thr	Gly	Arg	Leu 180	Phe	Glu	Leu	Asp	Gly 185	Leu	Lys	Val	Tyr	Pro 190	Ile	Asp
His	Gly	Pro 195	Trp	Gly	Glu	Asp	Glu 200	Glu	Trp	Thr	Asp	Lys 205	Ala	Arg	Arg
Val	Ile 210	Met	Glu	Arg	Ile	Gly 215	Leu	Ala	Thr	Ala	Gly 220	Glu	Pro	Tyr	His
Asp 225	Ile	Arg	Phe	Asn	Leu 230	Met	Ala	Val	Val	Pro 235	Asp	Arg	Arg	Ile	Lys 240
Tyr	Glu	Ala	Arg	Leu 245	His	Val	Leu	Lys	Val 250	Asn	Arg	Gln	Thr	Val 255	Leu
Glu	Ala	Leu	Gln 260	Gln	Leu	Ile	Arg	Val 265	Thr	Gln	Pro	Glu	Leu 270	Ile	Gln
Thr	His	Lys 275	Ser	Gln	Glu	Ser	Gln 280	Leu	Pro	Glu	Glu	Ser 285	Lys	Ser	Ala
Ser	Asn 290	Lys	Ser	Pro	Leu	Val 295	Leu	Glu	Ala	Asn	Arg 300	Ala	Pro	Ala	Ala
Ser 305	Glu	Gly	Asn	His	Thr 310	Asp	Gly	Ala	Glu	Glu 315	Ala	Ala	Gly	Ser	Cys 320
Ala	Gln	Ala	Pro	Ser 325	His	Ser	Pro	Pro	Asn 330	Lys	Pro	Lys	Leu	Val 335	Val
Lys	Pro	Pro	Gly 340	Ser	Ser	Leu	Asn	Gly 345	Val	His	Pro	Asn	Pro 350	Thr	Pro
Ile	Val	Gln 355	Arg	Leu	Pro	Ala	Phe 360	Leu	Asp	Asn	His	Asn 365	Tyr	Ala	Lys

58

Ser	Pro	Met	Gln	Glu	Glu	Glu	Asp	Leu	Ala	Ala	Gly	Val	Gly	Arg	Ser
	370					375					380				
Arg	Val	Pro	Val	Arg	Pro	Pro	Gln	Gln	Tyr	Ser	Asp	Asp	Glu	Asp	Asp
385					390					395					400
Tyr	Glu	Asp	Asp	Glu	Glu	Asp	Asp	Val	Gln	Asn	Thr	Asn	Ser	Ala	Leu
				405					410					415	
Arg	Tyr	Lys	Gly	Lys	Gly	Thr	Gly	Lys	Pro	Gly	Ala	Leu	Ser	Gly	Ser
			420					425					430		
Ala	Asp	Gly	Gln	Leu	Ser	Val	Leu	Gln	Pro	Asn	Thr	Ile	Asn	Val	Leu
		435					440					445			
Ala	Glu	Lys	Leu	Lys	Glu	Ser	Gln	Lys	Asp	Leu	Ser	Ile	Pro	Leu	Ser
	450					455					460				
Ile	Lys	Thr	Ser	Ser	Gly	Ala	Gly	Ser	Pro	Ala	Val	Ala	Val	Pro	Thr
465					470					475					480
His	Ser	Gln	Pro	Ser	Pro	Thr	Pro	Ser	Asn	Glu	Ser	Thr	Asp	Thr	Ala
				485					490					495	
Ser	Glu	Ile	Gly	Ser	Ala	Phe	Asn	Ser	Pro	Leu	Arg	Ser	Pro	Ile	Arg
			500					505					510		
Ser	Ala	Asn	Pro	Thr	Arg	Pro	Ser	Ser	Pro	Val	Thr	Ser	His	Ile	Ser
		515					520					525			
Lys	Val	Leu	Phe	Gly	Glu	Asp	Asp	Ser	Leu	Leu	Arg	Val	Asp	Cys	Ile
	530					535					540				
Arg	Tyr	Asn	Arg	Ala	Val	Arg	Asp	Leu	Gly	Pro	Val	Ile	Ser	Thr	Gly
545					550					555					560
Leu	Leu	His	Leu	Ala	Glu	Asp	Gly	Val	Leu	Ser	Pro	Leu	Ala	Leu	Thr
				565					570					575	
Glu	Gly	Gly	Lys	Gly	Ser	Ser	Pro	Ser	Ile	Arg	Pro	Ile	Gln	Gly	Ser
			580					585					590		
Gln	Gly	Ser	Ser	Ser	Pro	Val	Glu	Lys	Glu	Val	Val	Glu	Ala	Thr	Asp
		595					600					605			
Ser	Arg	Glu	Lys	Thr	Gly	Met	Val	Arg	Pro	Gly	Glu	Pro	Leu	Ser	Gly
	610					615					620				
Glu	Lys	Tyr	Ser	Pro	Lys	Glu	Leu	Leu	Ala	Leu	Leu	Lys	Cys	Val	Glu
625					630					635					640
Ala	Glu	Ile	Ala	Asn	Tyr	Glu	Ala	Cys	Leu	Lys	Glu	Glu	Val	Glu	Lys
				645					650					655	
Arg	Lys	Lys	Phe	Lys	Ile	Asp	Asp	Gln	Arg	Arg	Thr	His	Asn	Tyr	Asp
			660					665					670		
Glu	Phe	Ile	Cys	Thr	Phe	Ile	Ser	Met	Leu	Ala	Gln	Glu	Gly	Met	Leu
	675						680					685			
Ala	Asn	Leu	Val	Glu	Gln	Asn	Ile	Ser	Val	Arg	Arg	Arg	Gln	Gly	Val
	690					695					700				

59

Ser Ile Gly Arg Leu His Lys Gln Arg Lys Pro Asp Arg Arg Lys Arg
 705 710 715 720

Ser Arg Pro Tyr Lys Ala Lys Arg Gln
 725

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 100 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Met Asp Leu Ser Ala Leu Arg Val Glu Glu Val Gln Asn Val Ile Asn
 1 5 10 15

Ala Met Gln Lys Ile Leu Glu Cys Pro Ile Cys Leu Glu Leu Ile Lys
 20 25 30

Glu Pro Val Ser Thr Lys Cys Asp His Ile Phe Cys Lys Phe Cys Met
 35 40 45

Leu Lys Leu Leu Asn Gln Lys Lys Gly Pro Ser Gln Cys Pro Leu Cys
 50 55 60

Lys Asn Asp Ile Thr Lys Arg Ser Leu Gln Glu Ser Thr Arg Phe Ser
 65 70 75 80

Gln Leu Val Glu Glu Leu Leu Lys Ile Ile Cys Ala Phe Gln Leu Asp
 85 90 95

Thr Gly Leu Glu
 100

(2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 100 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met Asp Leu Ser Ala Val Gln Ile Gln Glu Val Gln Asn Val Leu His
 1 5 10 15

Ala Met Gln Lys Ile Leu Glu Cys Pro Ile Cys Leu Glu Leu Ile Lys
 20 25 30

Glu Pro Val Ser Thr Lys Cys Asp His Ile Phe Cys Lys Phe Cys Met
 35 40 45

60

Leu	Lys	Leu	Leu	Asn	Gln	Lys	Lys	Gly	Pro	Ser	Gln	Cys	Pro	Leu	Cys
	50					55					60				
Lys	Asn	Glu	Ile	Thr	Lys	Arg	Ser	Leu	Gln	Gly	Ser	Thr	Arg	Phe	Ser
65					70					75					80
Gln	Leu	Ala	Glu	Glu	Leu	Leu	Arg	Ile	Met	Ala	Ala	Phe	Glu	Leu	Asp
				85					90					95	
Thr	Gly	Met	Gln												
			100												

(2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 100 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Met	Ala	Ser	Ser	Val	Leu	Glu	Met	Ile	Lys	Glu	Glu	Val	Thr	Cys	Pro
1				5					10					15	
Ile	Cys	Leu	Glu	Leu	Leu	Lys	Glu	Pro	Val	Ser	Ala	Asp	Cys	Asn	His
			20					25					30		
Ser	Phe	Cys	Arg	Ala	Cys	Ile	Thr	Leu	Asn	Tyr	Glu	Ser	Asn	Arg	Asn
		35					40					45			
Thr	Asp	Gly	Lys	Gly	Asn	Cys	Pro	Val	Cys	Arg	Val	Pro	Tyr	Pro	Phe
	50					55					60				
Gly	Asn	Leu	Arg	Pro	Asn	Leu	His	Val	Ala	Asn	Ile	Val	Glu	Arg	Leu
65					70					75					80
Lys	Gly	Phe	Lys	Ser	Ile	Pro	Glu	Glu	Glu	Gln	Lys	Val	Asn	Ile	Cys
				85					90					95	
Ala	Gln	His	Gly												
			100												

(2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 262 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Lys	Lys	Glu	Ile	Trp	Asn	Ser	Asp	Pro	Arg	Gly	His	Glu	Gly	Pro	Gln
1				5					10					15	

61

Pro	Ser	Pro	Thr 20	Pro	Ser	Asn	Glu	Ser 25	Thr	Asp	Thr	Ala	Ser 30	Glu	Ile
Gly	Ser	Ala 35	Phe	Asn	Ser	Pro	Leu 40	Arg	Ser	Pro	Ile	Arg 45	Ser	Ala	Asn
Pro	Thr 50	Arg	Pro	Ser	Ser	Pro 55	Val	Thr	Ser	His	Ile 60	Ser	Lys	Val	Leu
Phe 65	Gly	Glu	Asp	Asp	Ser 70	Leu	Leu	Arg	Val	Asp 75	Cys	Ile	Arg	Tyr	Asn 80
Arg	Ala	Val	Arg	Asp 85	Leu	Gly	Pro	Val	Ile 90	Ser	Thr	Gly	Leu	Leu 95	His
Leu	Ala	Glu	Asp 100	Gly	Val	Leu	Ser	Pro 105	Leu	Ala	Leu	Thr	Glu 110	Gly	Gly
Lys	Gly	Ser 115	Ser	Pro	Ser	Ile	Arg 120	Pro	Ile	Gln	Gly	Ser 125	Gln	Gly	Ser
Ser	Ser 130	Pro	Val	Glu	Lys	Glu 135	Val	Val	Glu	Ala	Thr 140	Asp	Ser	Arg	Glu
Lys 145	Thr	Gly	Met	Val	Arg 150	Ser	Gly	Glu	Pro	Leu 155	Ser	Gly	Glu	Lys	Tyr 160
Ser	Pro	Lys	Glu	Leu 165	Leu	Ala	Leu	Leu	Lys 170	Cys	Val	Glu	Ala	Glu 175	Ile
Ala	Asn	Tyr	Glu 180	Ala	Cys	Leu	Lys	Glu 185	Glu	Val	Glu	Lys	Arg 190	Lys	Lys
Phe	Lys	Ile 195	Asp	Asp	Gln	Arg	Arg 200	Thr	His	Asn	Tyr	Asp 205	Glu	Phe	Ile
Cys	Thr 210	Phe	Ile	Ser	Met	Leu 215	Ala	Gln	Glu	Gly	Met 220	Leu	Ala	Asn	Leu
Val 225	Glu	Gln	Asn	Ile	Ser 230	Val	Arg	Arg	Arg	Gln 235	Gly	Val	Ser	Ile	Gly 240
Arg	Leu	His	Lys	Gln 245	Arg	Lys	Pro	Asp	Arg 250	Arg	Lys	Arg	Ser	Arg 255	Pro
Tyr	Lys	Ala	Lys 260	Arg	Gln										

(2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 188 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Gly	Gly	Ile	Asp	Trp	Ile	Pro	Gly	Tyr	Arg	Ala	Gln	Ile	Arg	Arg	Pro
1				5					10					15	

62

Ser	Ser	Pro	Val	Thr	Ser	His	Ile	Ser	Lys	Val	Leu	Phe	Gly	Glu	Asp
			20					25					30		
Asp	Ser	Leu	Leu	Arg	Val	Asp	Cys	Ile	Arg	Tyr	Asn	Arg	Ala	Val	Arg
		35					40					45			
Asp	Leu	Gly	Pro	Val	Ile	Ser	Thr	Gly	Leu	Leu	His	Leu	Ala	Glu	Asp
	50					55					60				
Gly	Val	Leu	Ser	Pro	Leu	Ala	Leu	Thr	Glu	Gly	Gly	Lys	Gly	Ser	Ser
65					70					75				80	
Pro	Ser	Thr	Arg	Ser	Ser	Gln	Gly	Ser	Gln	Gly	Ser	Ser	Gly	Leu	Glu
				85					90					95	
Glu	Lys	Glu	Val	Val	Glu	Val	Thr	Glu	Ser	Arg	Asp	Lys	Pro	Gly	Leu
			100					105					110		
Asn	Arg	Ser	Ser	Glu	Pro	Leu	Ser	Gly	Glu	Lys	Tyr	Ser	Pro	Lys	Ile
		115					120					125			
Asp	Asp	Gln	Arg	Arg	Thr	His	Asn	Tyr	Asp	Glu	Phe	Ile	Cys	Thr	Phe
	130					135					140				
Ile	Ser	Met	Leu	Ala	Gln	Glu	Gly	Met	Leu	Ala	Asn	Leu	Val	Glu	Gln
145					150					155				160	
Asn	Ile	Ser	Val	Arg	Arg	Arg	Gln	Gly	Val	Ser	Ile	Gly	Arg	Leu	His
				165					170					175	
Lys	Gln	Arg	Lys	Pro	Asp	Arg	Arg	Met	Ser	Gly	Arg				
			180					185							

(2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 161 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Gly	Gly	Ile	Asp	Trp	Ile	Pro	Gly	Val	Arg	Ala	Gln	Ile	Arg	Pro	Ile
1				5					10					15	
Ser	Ser	Ser	Ser	Pro	Ser	Thr	Arg	Ser	Ser	Gln	Gly	Ser	Gln	Gly	Ser
			20					25					30		
Ser	Gly	Leu	Glu	Glu	Lys	Glu	Val	Val	Glu	Val	Thr	Glu	Ser	Arg	Asp
		35					40				45				
Lys	Pro	Gly	Leu	Asn	Arg	Ser	Ser	Glu	Pro	Leu	Ser	Gly	Glu	Lys	Tyr
	50					55					60				
Ser	Pro	Lys	Glu	Leu	Leu	Ala	Leu	Leu	Lys	Cys	Ala	Glu	Ala	Glu	Ile
65					70					75				80	
Ala	Asn	Tyr	Glu	Ala	Cys	Leu	Lys	Glu	Glu	Val	Glu	Lys	Arg	Lys	Lys
				85					90					95	

[illegible]

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH:	149	amino	acids
(B) TYPE:	amino	acid	
(C) STRANDEDNESS:			
(D) TOPOLOGY:	unknown		

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

[illegible]

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 80 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

ATGGACCTGT	CTGCTCTGCG	TGTTGAAGAA	GTTCAAAACG	TTATCAACGC	TATGCAAAAG	60
ATCCTGGAAT	GTCCAATCTG					80

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 101 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GGTTCAGCAG	CTTCAGCATA	CAGAACTTAC	AGAAGATGTG	GTCACACTTA	GTGGAAACTG	60
GTTCCCTGAT	CAGTTCAGA	CAGATTGGAC	ATTCCAGGAT	C		101

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 102 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

GTATGCTGAA	GCTGCTGAAC	CAAAGAAGG	GTCCATCTCA	ATGTCCACTG	TGTAAGAACG	60
ACATCACTAA	GCGTTCTCTG	CAAGAATCTA	CTCGTTTCTC	TC		102

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 81 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: other nucleic acid

65

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

TTCCAGACCA	GTGTCCAGCT	GGAAAGCACA	GATGATCTTC	AGCAGTTCTT	CAACCAGTTG	60
AGAGAAACGA	GTAGATTCTT	G				81

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 31 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

GCTAGAATTC	ACCATGGACC	TGTCTGCTCT	G	31
------------	------------	------------	---	----

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 28 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

GCTAGTCGAC	TTCCAGACCA	GTGTCCAG	28
------------	------------	----------	----

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 247 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

Met	Gly	Lys	Lys	Ile	Met	Thr	Asp	Ala	Gly	Ser	Trp	Cys	Leu	Ile	Glu
1				5					10					15	
Ser	Asp	Pro	Gly	Val	Phe	Thr	Glu	Met	Leu	Arg	Gly	Phe	Gly	Val	Asp
			20					25					30		
Gly	Leu	Gln	Val	Glu	Glu	Leu	Tyr	Ser	Leu	Asp	Asp	Asp	Lys	Ala	Met
		35					40				45				
Thr	Arg	Pro	Thr	Tyr	Gly	Leu	Ile	Phe	Leu	Phe	Lys	Trp	Arg	Gln	Gly
	50					55					60				
Asp	Glu	Thr	Thr	Gly	Ile	Pro	Ser	Asp	Lys	Gln	Asn	Ile	Phe	Phe	Ala
65					70					75					80

66

His	Gln	Thr	Ile	Gln 85	Asn	Ala	Cys	Ala	Thr 90	Gln	Ala	Leu	Ile	Asn 95	Leu
Leu	Met	Asn	Val 100	Glu	Asp	Thr	Asp	Val 105	Lys	Leu	Gly	Asn	Ile 110	Leu	Asn
Gln	Tyr	Lys 115	Glu	Phe	Ala	Ile	Asp 120	Leu	Asp	Pro	Asn	Thr 125	Arg	Gly	His
Cys	Leu 130	Ser	Asn	Ser	Glu	Glu 135	Ile	Arg	Thr	Val	His 140	Asn	Ser	Phe	Ser
Arg 145	Gln	Thr	Leu	Phe	Glu 150	Leu	Asp	Ile	Lys	Gly 155	Gly	Glu	Ser	Glu	Asp 160
Asn	Tyr	His	Phe	Val 165	Thr	Tyr	Val	Pro	Ile 170	Gly	Asn	Lys	Val	Tyr 175	Glu
Leu	Asp	Gly	Leu 180	Arg	Glu	Leu	Pro	Leu 185	Glu	Val	Ala	Glu	Phe 190	Gln	Lys
Glu	Gln	Asp 195	Trp	Ile	Glu	Ala	Ile 200	Lys	Pro	Val	Ile	Gln 205	Gln	Arg	Met
Gln	Lys 210	Tyr	Ser	Glu	Gly	Glu 215	Ile	Thr	Phe	Asn	Leu 220	Met	Ala	Leu	Val
Pro 225	Asn	Arg	Lys	Gln	Lys 230	Leu	Gln	Glu	Met	Met 235	Glu	Asn	Leu	Ile	Gln 240
Ala	Asn	Glu	Asn	Asn 245	Glu	Leu									

(2) INFORMATION FOR SEQ ID NO:17:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 223 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

Met 1	Gln	Leu	Lys	Pro 5	Met	Glu	Ile	Asn	Pro 10	Glu	Met	Leu	Asn	Lys 15	Val
Leu	Ser	Arg	Leu 20	Gly	Val	Ala	Gly	Gln 25	Trp	Arg	Phe	Val	Asp 30	Val	Leu
Gly	Leu	Glu 35	Glu	Glu	Ser	Leu	Gly 40	Ser	Val	Pro	Ala	Pro 45	Ala	Cys	Ala
Leu 50	Leu	Leu	Leu	Phe	Pro	Leu 55	Thr	Ala	Gln	His	Glu 60	Asn	Phe	Arg	Lys
Lys 65	Gln	Ile	Glu	Glu	Leu 70	Lys	Gly	Gln	Glu	Val 75	Ser	Pro	Lys	Val	Tyr 80
Phe	Met	Lys	Gln	Thr 85	Ile	Gly	Asn	Ser	Cys 90	Gly	Thr	Ile	Gly	Leu 95	Ile

67

His	Ala	Val	Ala 100	Asn	Asn	Gln	Asp	Lys 105	Leu	Gly	Phe	Glu	Asp 110	Gly	Ser
Val	Leu	Lys 115	Gln	Phe	Leu	Ser	Glu 120	Thr	Glu	Lys	Met	Ser 125	Pro	Glu	Asp
Arg	Ala 130	Lys	Cys	Phe	Glu	Lys 135	Asn	Glu	Ala	Ile	Gln 140	Ala	Ala	His	Asp
Ala 145	Val	Ala	Gln	Glu	Gly 150	Gln	Cys	Arg	Val	Asp 155	Asp	Lys	Val	Asn	Phe 160
His	Phe	Ile	Leu	Phe 165	Asn	Asn	Val	Asp	Gly 170	His	Leu	Tyr	Glu	Leu 175	Asp
Gly	Arg	Met	Pro 180	Phe	Pro	Val	Asn	His 185	Gly	Ala	Ser	Ser	Glu 190	Asp	Thr
Leu	Leu	Lys 195	Asp	Ala	Ala	Lys	Val 200	Cys	Arg	Glu	Phe	Thr 205	Glu	Arg	Glu
Gln	Gly 210	Glu	Val	Arg	Phe	Ser 215	Ala	Val	Ala	Leu	Cys 220	Lys	Ala	Ala	

(2) INFORMATION FOR SEQ ID NO:18:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 230 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

Met 1	Glu	Gly	Gln	Arg 5	Trp	Leu	Pro	Leu	Glu 10	Ala	Asn	Pro	Glu	Val 15	Thr
Asn	Gln	Phe	Leu 20	Lys	Gln	Leu	Gly	Leu 25	His	Pro	Asn	Trp	Gln 30	Phe	Val
Asp	Val	Tyr 35	Gly	Met	Asp	Pro	Glu 40	Leu	Leu	Ser	Met	Val 45	Pro	Arg	Pro
Val	Cys 50	Ala	Val	Leu	Leu	Leu 55	Phe	Pro	Ile	Thr	Glu 60	Lys	Tyr	Glu	Val
Phe 65	Arg	Thr	Glu	Glu 70	Glu	Lys	Ile	Lys	Ser 75	Gln	Gly	Gln	Asp	Val 80	
Thr	Ser	Ser	Val	Tyr 85	Phe	Met	Lys	Gln	Thr 90	Ile	Ser	Asn	Ala	Cys 95	Gly
Thr	Ile	Gly	Leu 100	Ile	His	Ala	Ile	Ala 105	Asn	Asn	Lys	Asp	Lys 110	Met	His
Phe	Glu	Ser 115	Gly	Ser	Thr	Leu	Lys 120	Lys	Phe	Leu	Glu	Glu 125	Ser	Val	Ser
Met 130	Ser	Pro	Glu	Glu	Arg	Ala 135	Arg	Tyr	Leu	Glu	Asn 140	Tyr	Asp	Ala	Ile

68

Arg 145	Val	Thr	His	Glu	Thr 150	Ser	Ala	His	Glu	Gly 155	Gln	Thr	Glu	Ala	Pro 160
Ser	Ile	Asp	Glu	Lys 165	Val	Asp	Leu	His	Phe 170	Ile	Ala	Leu	Val	His 175	Val
Asp	Gly	His	Leu 180	Tyr	Glu	Leu	Asp	Gly 185	Arg	Lys	Pro	Phe	Pro 190	Ile	Asn
His	Gly	Glu 195	Thr	Ser	Asp	Glu	Thr 200	Leu	Leu	Glu	Asp	Ala 205	Ile	Glu	Val
Cys	Lys 210	Lys	Phe	Met	Glu	Arg 215	Asp	Pro	Asp	Glu	Leu 220	Arg	Phe	Asn	Ala
Ile 225	Ala	Leu	Ser	Ala	Ala 230										

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 43 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

Leu 1	Leu	Arg	Cys	Ser 5	Arg	Cys	Thr	Asn	Ile 10	Leu	Arg	Glu	Pro	Val 15	Cys
Leu	Gly	Gly	Cys 20	Glu	His	Ile	Phe	Cys 25	Ser	Asn	Cys	Val	Ser 30	Asp	Cys
Ile	Gly	Thr 35	Gly	Cys	Pro	Val	Cys 40	Tyr	Thr	Pro					

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 326 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

Met 1	Gly	Lys	Lys	Ile 5	Met	Thr	Asp	Ala	Gly 10	Ser	Trp	Cys	Leu	Ile 15	Glu
Ser	Asp	Pro	Gly 20	Val	Phe	Thr	Glu	Met 25	Leu	Arg	Gly	Phe	Gly 30	Val	Asp
Gly	Leu	Gln 35	Val	Glu	Glu	Leu	Tyr 40	Ser	Leu	Asp	Asp	Asp 45	Lys	Ala	Met

69

Thr	Arg	Pro	Thr	Tyr	Gly	Leu	Ile	Phe	Leu	Phe	Lys	Trp	Arg	Gln	Gly
	50					55					60				
Asp	Glu	Thr	Thr	Gly	Ile	Pro	Ser	Asp	Lys	Gln	Asn	Ile	Phe	Phe	Ala
65					70					75					80
His	Gln	Thr	Ile	Gln	Asn	Ala	Cys	Ala	Thr	Gln	Ala	Leu	Ile	Asn	Leu
				85					90					95	
Leu	Met	Asn	Val	Glu	Asp	Thr	Asp	Val	Lys	Leu	Gly	Asn	Ile	Leu	Asn
			100					105					110		
Gln	Tyr	Lys	Glu	Phe	Ala	Ile	Asp	Leu	Asp	Pro	Asn	Thr	Arg	Gly	His
		115					120					125			
Cys	Leu	Ser	Asn	Ser	Glu	Glu	Ile	Arg	Thr	Val	His	Asn	Ser	Phe	Ser
	130					135					140				
Arg	Gln	Thr	Leu	Phe	Glu	Leu	Asp	Ile	Lys	Gly	Gly	Glu	Ser	Glu	Asp
145					150					155					160
Asn	Tyr	His	Phe	Val	Thr	Tyr	Val	Pro	Ile	Gly	Asn	Lys	Val	Tyr	Glu
				165					170					175	
Leu	Asp	Gly	Leu	Arg	Glu	Leu	Pro	Leu	Glu	Val	Ala	Glu	Phe	Gln	Lys
			180					185					190		
Glu	Gln	Asp	Trp	Ile	Glu	Ala	Ile	Lys	Pro	Val	Ile	Gln	Gln	Arg	Met
		195					200					205			
Gln	Lys	Tyr	Ser	Glu	Gly	Glu	Ile	Thr	Phe	Asn	Leu	Met	Ala	Leu	Val
	210					215					220				
Pro	Asn	Arg	Lys	Gln	Lys	Leu	Gln	Glu	Met	Met	Glu	Asn	Leu	Ile	Gln
225					230					235					240
Ala	Asn	Glu	Asn	Asn	Glu	Leu	Glu	Glu	Gln	Ile	Ala	Asp	Leu	Asn	Lys
				245					250					255	
Ala	Ile	Ala	Asp	Glu	Asp	Tyr	Lys	Met	Glu	Met	Tyr	Arg	Lys	Glu	Asn
			260					265					270		
Asn	Arg	Arg	Arg	His	Asn	Tyr	Thr	Pro	Phe	Val	Ile	Glu	Leu	Met	Lys
		275					280					285			
Ile	Leu	Ala	Lys	Glu	Gly	Lys	Leu	Val	Gly	Leu	Val	Asp	Asn	Ala	Tyr
	290					295					300				
Gln	Ala	Ala	Lys	Glu	Lys	Ser	Lys	Leu	Asn	Thr	Asp	Ile	Thr	Lys	Leu
305					310					315					320
Glu	Leu	Lys	Arg	Lys	Gln										
				325											

(2) INFORMATION FOR SEQ ID NO:21:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 227 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: unknown

70

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

Met 1	Leu	Thr	Trp	Thr 5	Pro	Leu	Glu	Ser	Asn 10	Pro	Glu	Val	Leu	Thr 15	Lys
Tyr	Ile	His	Lys 20	Leu	Ala	Val	Ser	Pro 25	Ala	Trp	Ser	Val	Thr 30	Asp	Val
Ile	Gly	Leu 35	Glu	Asp	Asp	Thr	Leu 40	Glu	Trp	Ile	Pro	Arg 45	Pro	Val	Lys
Ala	Phe 50	Ile	Leu	Leu	Phe	Pro 55	Cys	Ser	Glu	Thr	Tyr 60	Glu	Lys	His	Arg
Thr 65	Glu	Glu	His	Asp	Arg 70	Ile	Lys	Glu	Val	Glu 75	Glu	Gln	His	Pro	Glu 80
Asp	Leu	Phe	Tyr	Met 85	Arg	Gln	Phe	Thr	His 90	Asn	Ala	Cys	Gly	Thr 95	Val
Ala	Leu	Ile	His 100	Ser	Val	Ala	Asn	Asn 105	Lys	Glu	Val	Asp	Ile 110	Asp	Arg
Gly	Val	Leu 115	Lys	Asp	Phe	Leu	Glu 120	Lys	Thr	Ala	Ser	Leu 125	Ser	Pro	Glu
Glu	Arg 130	Gly	Arg	Ala	Leu	Glu 135	Lys	Asp	Glu	Lys	Phe 140	Thr	Ala	Asp	His
Glu 145	Ala	Leu	Ala	Gln	Glu 150	Gly	Gln	Thr	Asn	Ala 155	Ala	Asn	His	Glu	Lys 160
Val	Ile	His	His	Phe 165	Ile	Ala	Leu	Val	Asn 170	Lys	Glu	Gly	Thr	Leu 175	Tyr
Glu	Leu	Asp	Gly 180	Arg	Lys	Ser	Phe	Pro 185	Ile	Lys	His	Gly	Pro 190	Thr	Ser
Glu	Glu	Thr 195	Phe	Val	Lys	Asp	Ala 200	Ala	Lys	Val	Cys	Lys 205	Glu	Phe	Met
Ala	Arg 210	Asp	Pro	Asn	Glu	Val 215	Arg	Phe	Thr	Val	Leu 220	Ala	Leu	Thr	Ala
Ala 225	Gln	Gln													

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 236 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

71

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

Met	Ser	Gly	Glu	Asn	Arg	Ala	Val	Val	Pro	Ile	Glu	Ser	Asn	Pro	Glu
1				5					10					15	
Val	Phe	Thr	Asn	Phe	Ala	His	Lys	Leu	Gly	Leu	Lys	Asn	Glu	Trp	Ala
			20					25					30		
Tyr	Phe	Asp	Ile	Tyr	Ser	Leu	Thr	Glu	Pro	Glu	Leu	Leu	Ala	Phe	Leu
		35					40					45			
Pro	Arg	Pro	Val	Lys	Ala	Ile	Val	Leu	Leu	Phe	Pro	Ile	Asn	Glu	Asp
	50					55					60				
Arg	Lys	Ser	Ser	Thr	Ser	Gln	Gln	Ile	Thr	Ser	Ser	Tyr	Asp	Val	Ile
65					70					75					80
Trp	Phe	Lys	Gln	Ser	Val	Lys	Asn	Ala	Cys	Gly	Leu	Tyr	Ala	Ile	Leu
				85					90					95	
His	Ser	Leu	Ser	Asn	Asn	Gln	Ser	Leu	Leu	Glu	Pro	Gly	Ser	Asp	Leu
			100					105					110		
Asp	Asn	Phe	Leu	Lys	Ser	Gln	Ser	Asp	Thr	Ser	Ser	Ser	Lys	Asn	Arg
		115					120					125			
Phe	Asp	Asp	Val	Thr	Thr	Asp	Gln	Phe	Val	Leu	Asn	Val	Ile	Lys	Glu
	130					135					140				
Asn	Val	Gln	Thr	Phe	Ser	Thr	Gly	Gln	Ser	Glu	Ala	Pro	Glu	Ala	Thr
145					150					155					160
Ala	Asp	Thr	Asn	Leu	His	Tyr	Ile	Thr	Tyr	Val	Glu	Glu	Asn	Gly	Gly
				165					170					175	
Ile	Phe	Glu	Leu	Asp	Gly	Arg	Asn	Leu	Ser	Gly	Pro	Leu	Tyr	Leu	Gly
			180					185					190		
Lys	Ser	Asp	Pro	Thr	Ala	Thr	Asp	Leu	Ile	Glu	Gln	Glu	Leu	Val	Arg
		195					200					205			
Val	Arg	Val	Ala	Ser	Tyr	Met	Glu	Asn	Ala	Asn	Glu	Glu	Asp	Val	Leu
	210					215					220				
Asn	Phe	Ala	Met	Leu	Gly	Leu	Gly	Pro	Asn	Trp	Glu				
225					230					235					

(2) INFORMATION FOR SEQ ID NO:23:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 230 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

Met	Glu	Gly	Gln	Arg	Trp	Leu	Pro	Leu	Glu	Ala	Asn	Pro	Glu	Val	Thr
1				5					10					15	

72

Asn	Gln	Phe	Leu 20	Lys	Gln	Leu	Gly	Leu 25	His	Pro	Asn	Trp	Gln 30	Phe	Val
Asp	Val	Tyr 35	Gly	Met	Asp	Pro	Glu 40	Leu	Leu	Ser	Met	Val 45	Pro	Arg	Pro
Val	Cys 50	Ala	Val	Leu	Leu	Leu 55	Phe	Pro	Ile	Thr	Glu 60	Lys	Tyr	Glu	Val
Phe 65	Arg	Thr	Glu	Glu	Glu 70	Glu	Lys	Ile	Lys	Ser 75	Gln	Gly	Gln	Asp	Val 80
Thr	Ser	Ser	Val	Tyr 85	Phe	Met	Lys	Gln	Thr 90	Ile	Ser	Asn	Ala	Cys 95	Gly
Thr	Ile	Gly	Leu 100	Ile	His	Ala	Ile	Ala 105	Asn	Asn	Lys	Asp	Lys 110	Met	His
Phe	Glu	Ser 115	Gly	Ser	Thr	Leu	Lys 120	Lys	Phe	Leu	Glu	Glu 125	Ser	Val	Ser
Met	Ser 130	Pro	Glu	Glu	Arg	Ala 135	Arg	Tyr	Leu	Glu	Asn 140	Tyr	Asp	Ala	Ile
Arg 145	Val	Thr	His	Glu	Thr 150	Ser	Ala	His	Glu	Gly 155	Gln	Thr	Glu	Ala	Pro 160
Ser	Ile	Asp	Glu	Lys 165	Val	Asp	Leu	His	Phe 170	Ile	Ala	Leu	Val	His 175	Val
Asp	Gly	His	Leu 180	Tyr	Glu	Leu	Asp	Gly 185	Arg	Lys	Pro	Phe	Pro 190	Ile	Asn
His	Gly	Glu 195	Thr	Ser	Asp	Glu	Thr 200	Leu	Leu	Glu	Asp	Ala 205	Ile	Glu	Val
Cys	Lys 210	Lys	Phe	Met	Glu	Arg 215	Asp	Pro	Asp	Glu	Leu 220	Arg	Phe	Asn	Ala
Ile 225	Ala	Leu	Ser	Ala	Ala 230										

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 223 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

Met 1	Gln	Leu	Lys	Pro 5	Met	Glu	Ile	Asn	Pro 10	Glu	Met	Leu	Asn	Lys 15	Val
Leu	Ser	Arg	Leu 20	Gly	Val	Ala	Gly	Gln 25	Trp	Arg	Phe	Val	Asp 30	Val	Leu
Gly	Leu	Glu 35	Glu	Glu	Ser	Leu	Gly 40	Ser	Val	Pro	Ala	Pro 45	Ala	Cys	Ala

73

Leu 50	Leu	Leu	Leu	Phe	Pro	Leu 55	Thr	Ala	Gln	His	Glu 60	Asn	Phe	Arg	Lys
Lys 65	Gln	Ile	Glu	Glu	Leu 70	Lys	Gly	Gln	Glu	Val 75	Ser	Pro	Lys	Val	Tyr 80
Phe	Met	Lys	Gln	Thr 85	Ile	Gly	Asn	Ser	Cys 90	Gly	Thr	Ile	Gly	Leu 95	Ile
His	Ala	Val	Ala 100	Asn	Asn	Gln	Asp	Lys 105	Leu	Gly	Phe	Glu	Asp 110	Gly	Ser
Val	Leu	Lys 115	Gln	Phe	Leu	Ser	Glu 120	Thr	Glu	Lys	Met	Ser 125	Pro	Glu	Asp
Arg	Ala 130	Lys	Cys	Phe	Glu	Lys 135	Asn	Glu	Ala	Ile	Gln 140	Ala	Ala	His	Asp
Ala 145	Val	Ala	Gln	Glu	Gly 150	Gln	Cys	Arg	Val	Asp 155	Asp	Lys	Val	Asn	Phe 160
His	Phe	Ile	Leu	Phe 165	Asn	Asn	Val	Asp	Gly 170	His	Leu	Tyr	Glu	Leu 175	Asp
Gly	Arg	Met	Pro 180	Phe	Pro	Val	Asn	His 185	Gly	Ala	Ser	Ser	Glu 190	Asp	Thr
Leu	Leu	Lys 195	Asp	Ala	Ala	Lys	Val 200	Cys	Arg	Glu	Phe	Thr 205	Glu	Arg	Glu
Gln	Gly 210	Glu	Val	Arg	Phe	Ser 215	Ala	Val	Ala	Leu	Cys 220	Lys	Ala	Ala	

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

CCATCTCAAG GTCCACTGTG TAAG

24

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: other nucleic acid

(iv) ANTI-SENSE: YES

74

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:
CTTACACAGT GGACCTTGAG ATGG

24

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:
CAATGTCCAC TGGGTAAGAA CGACATC

27

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: other nucleic acid

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:
GATGTCGTTT TTACCCAGTG GACATTG

27

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 79 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:
GCATGGATCC TCAAACCTTG TGCAGGCAGG TACCCTGGTC AACAGGAGAC AGGTGGGAAA
CCAGGATCTT TTGCATAGC

60

79

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

75

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

CCGATGCCCT TGAATTGAC GAG

23

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

CGATGAATTC GAGCTAGCTT CTATC

25

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 32 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

GCATGAATTC TCAGCTCCGG CGCACTGAGA TG

32

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 33 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

GCATGAATTC TCAAGCCAGC ATGGATATGA AGG

33

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 32 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: other nucleic acid

76

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:
GCATGAATTC TCAGTCATCA ATCTTGAAC TC 32

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:
GCATGAATTC TCATGCAATC TCGGCTTCTA C 31

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 33 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:
GCATGGATCC CCAAGATTGA TGACCAGCGA AGG 33

(2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:
GCTGGCCAAC CCGGTGGAAC AG 22

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: other nucleic acid

(iv) ANTI-SENSE: YES

77

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:
CTGTTCCACC GGGTTGGCCA GC 22

(2) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 56 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:
CCTGTTATTA ACCCTCACTA AAGGGAAGGG TACCATGAAT AAGGGCTGGC TGGAGC 56

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:
GAAGCGGATG TCGTGGTAGG 20

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:
GATGTATATA ACTATCTATT CG 22

(2) INFORMATION FOR SEQ ID NO:42:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 34 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: other nucleic acid

78

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:
GCATAGATCT TCACCCCTGG CTGCCTTGGG TTGG

34

(2) INFORMATION FOR SEQ ID NO:43:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:
GAAGCGGATG TCGTGGTAGG

20

(2) INFORMATION FOR SEQ ID NO:44:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:
GATGTATATA ACTATCTATT CG

22

(2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:
CGTAGTCGAC TGTGAGCGCC AGGGGACTC

29

(2) INFORMATION FOR SEQ ID NO:46:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: other nucleic acid

79

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

CAACCCCACT CCCATTGTC

19

(2) INFORMATION FOR SEQ ID NO:47:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: other nucleic acid

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

GAGTTGGTGT TCTGCACGTC

20

WHAT IS CLAIMED IS:

1. A nucleic acid sequence encoding mammalian BRCA1 Associated Protein (BAP-1) or a fragment thereof, isolated from cellular materials with which it is naturally associated.
2. The nucleic acid sequence according to claim 1 which is selected from the group consisting of:
 - (a) SEQ ID NO:1;
 - (b) a sequence which hybridizes to (a) under stringent conditions;
 - (c) an allelic variant of (a) or (b); and
 - (d) a fragment of (a) or (b).
3. The sequence according to claim 2 wherein said fragment is selected from the group consisting of:
 - (a) open reading frame, nucleotides about 40 to about 2226 of SEQ ID NO:1;
 - (b) a region of acidity, nucleotides about 1225 to about 1263 of SEQ ID NO:1; and
 - (c) interactive domain, nucleotides about 1831 to about 2226 of SEQ ID NO:1.
4. The sequence according to claim 1 which encodes human BAP-1 SEQ ID NO:2 or a fragment thereof.
5. A mammalian BRCA1 associated protein (BAP-1) or a peptide fragment thereof.
6. The BAP-1 protein according to claim 5, said protein comprising an amino acid sequence selected from the group consisting of:
 - (a) human BAP-1, SEQ ID NO:2;
 - (b) fragment of (a);

- (c) analogues of (a) characterized by having at least about 85% homology with SEQ ID NO: 2; and
- (d) homologs of (a) characterized by having at least about 85% homology with SEQ ID NO: 2.

7. The BAP-1 protein according to claim 5, wherein said fragment is selected from the group consisting of:

- (a) amino acids about 656 to about 661 of SEQ ID NO:2;
- (b) amino acids about 717 to about 722 of SEQ ID NO:2;
- (c) amino acids about 396 to about 408 SEQ ID NO:2,
- (d) amino acids about 598 to about 729 of SEQ ID NO:2;
- (e) amino acids about 483 to about 576 of SEQ ID NO:2;
- (f) amino acids about 1 to about 214 of SEQ ID NO: 2;
- (g) amino acids about 1 to about 426 of SEQ ID NO: 2;
- (h) amino acids about 1 to about 352 of SEQ ID NO: 2;
- (i) amino acids about 1 to about 325 of SEQ ID NO: 2;
- (j) amino acids about 1 to about 313 of SEQ ID NO: 2; and
- (k) smaller fragments of (a) - (j) comprising about 8 amino acids.

8. A vector comprising a mammalian nucleic acid sequence encoding a BRCA1 associated protein (BAP-1) or peptide under the control of suitable regulatory sequences.

9. The vector according to claim 8, wherein said vector is a gene therapy vector.

10. A host cell transformed with the vector according to claim 8.

11. A method of recombinantly expressing BRCA1 associated protein (BAP-1) by culturing a recombinant host cell transformed with nucleic acid sequence encoding BAP-1 under conditions which permit expression of BAP-1.

12. A diagnostic reagent comprising a nucleic acid sequence selected from the group consisting of:

- (a) SEQ ID NO:1 and its complementary sequence;
- (b) a nucleotide sequence encoding amino acids 598 to 729 of SEQ ID NO: 2 and its complementary sequence;
- (c) a nucleic acid fragment of (a) or (b) comprising at least 15 nucleotides in length;
- (d) a sequence which hybridizes to (a), (b) or (c) under stringent conditions;
and a detectable label which is associated with said sequence.

13. An anti-BRCA1 associated protein (BAP-1) antibody.

14. The antibody according to claim 13, wherein said antibody binds to a peptide selected from the group consisting of

- (a) SEQ ID NO: 2;
- (b) amino acids 483 to 576 of SEQ ID NO: 2;
- (c) amino acids 598 to 729 of SEQ ID NO: 2;

and

(d) a fragment of (a), (b) or (c) comprising about 8 amino acids.

15. The antibody according to claim 13, selected from the group consisting of a chimeric antibody, a humanized antibody, a monoclonal antibody and a polyclonal antibody.

16. An anti-idiotypic antibody specific for the antibody of claim 13.

17. A diagnostic reagent comprising the antibody according to claim 13 and a detectable label.

18. A method of detecting a cancer involving BRCA1 comprising providing a biopsy sample from a patient suspected of having said cancer and incubating said sample in the presence of a diagnostic reagent according to claim 12 or 17.

19. A method of detecting a deficiency in BRCA1 associated protein (BAP-1) in a patient comprising providing a sample from a patient suspected of having said deficiency and performing the polymerase chain reaction using the diagnostic reagent according to claim 12.

20. A method of identifying compounds which specifically bind to BAP-1 or a fragment thereof, comprising the steps of contacting said BAP-1 or fragment with a test compound to permit binding of the test compound to BAP-1; and determining the amount of test compound which is bound to BAP-1.

21. The method according to claim 20 wherein said BAP-1 is immobilized on a solid support.

22. A compound identified by the method of claim 20.

1 / 1 1

FIG. 1A

hBRCA1:	1	MDLSAIRL	RVVE	VQNV	INAMQK	ILEC	PICL	LEI	*	*	*	*	*	*	IKEPVSTK	C	DHIFCKFCML	KLINQKK	----
BRCA1:	1	MDLSAVQIQE	VONV	LLHAMQK	ILEC	PICL	LEI								IKEPVSTK	C	DHIFCKFCML	KLINQKK	----
hrPT-1:									12	EVT	CPICL	LEI			IKEPVSAD	C	NHSFCRA	DT	LNYESNRNTD
hBARD1:									47	LP	CSRG	TNI			LRPEVCLGC		EHIFCSNOVS	DCI	-----

	hBRCA1:	hBRCA1:	hBRCA1:	hRPT-1:	hBARD1:
	GF	GF	GF	GKGN	GIG
	SO	SO	SO	CF	CE
	CH	CH	CH	CV	CV
	CK	CK	CK	CK	CK
	KN	KN	KN	KN	KN
	*	*	*	*	*
	DITKRS	EITKRS	EITKRS	PYPFGN	LHVANIV
	SLQES	SLQGS	SLQAEEL	LRPM	ERL
	TRFSQ	TRGSQ	TRGSQ	LAEEEL	KGFKS
	VEEL	VEEL	VEEL	LAEEEL	SIPEEQ
	LKIICAF	LRINAA	LRINAA	FOLD	KGFKS
	FOLD	FOLD	FOLD	KGFKS	SIPEEQ
	TGLE	TGHQ	TGHQ	KGFKS	SIPEEQ
	100	100	100	100	100
	100	100	100	100	100
	100	100	100	100	100
	100	100	100	100	100
	89	89	89	89	89

2 / 1 1

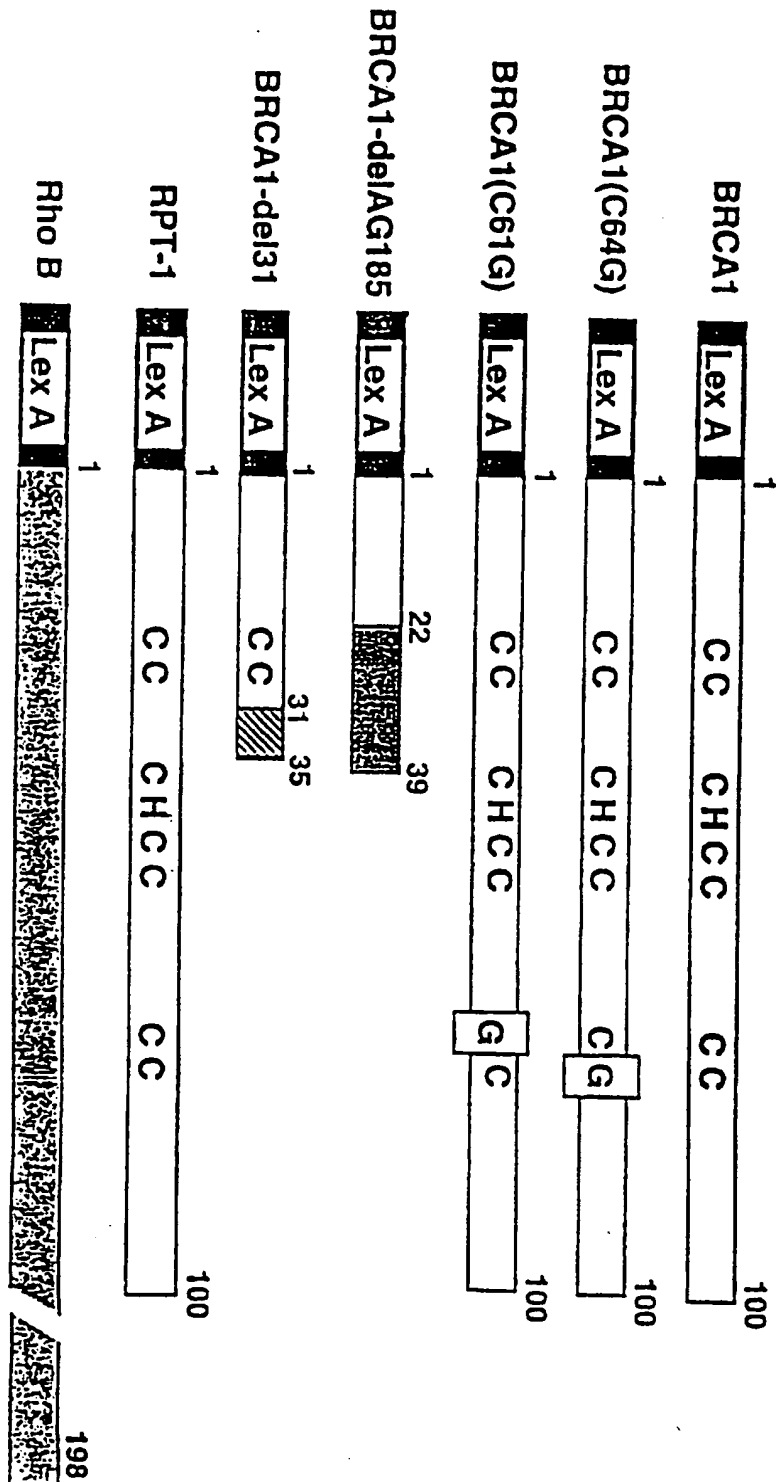


FIG. 1B

hBAP-1(483-729)	kkeiwnsdPr9he	----	GPQSP	TPSNE	STDA	SEIG	S	AFNS	PLRSP	I
mBAP-1(518-(del)-718)	ggidwlp9yraqir	-----								
mBAP-1(581-720)	ggidwlp9vraqirpis	-----								
mBAP-1(596-721)	ggidwlp9yraqirpis	-----								
hBAP-1(483-729)	RSANPTRPSSPVTSHISKVLFGEDDSLLRVDCIRYNRAVRDLGPVJSTG									
mBAP-1(518-(del)-718)	-----	RPSSPVTSHISKVLFGEDDSLLRVDCIRYNRAVRDLGPVJSTG								
mBAP-1(581-720)	-----									
mBAP-1(596-721)	-----									
hBAP-1(483-729)	LLHLAEDGVLSPALTEGGKQSSPS	IRPI	QGSQGS	SSPV	EKEV	VEAT	DS			
mBAP-1(518-(del)-718)	LLHLAEDGVLSPALTEGGKQSSPS	TRSS	QGSQGS	SGLE	EKEV	VEV	TES			
mBAP-1(581-720)	-----	TRSS	QGSQGS	SGLE	EKEV	VEV	TES			
mBAP-1(596-721)	-----	-----	-----	-----	-----	-----	-----	SGLE	EKEV	VEV
hBAP-1(483-729)	REKTMVRS	GEPLSGEKYSPKELLALLKC	VEAEI	ANYEAC	LKEE	VEE	KRK			
mBAP-1(518-(del)-718)	RDKPGLNRSSEPLSGEKYSP	-----	-----	-----	-----	-----	-----			
mBAP-1(581-720)	RDKPGLNRSSEPLSGEKYSPKELLALLKC	AEAEI	ANYEAC	LKEE	VEE	KRK				
mBAP-1(596-721)	RDKPGLNRSSEPLSGEKYSPKELLALLKC	VEAEI	ANYEAC	LKEE	VEE	KRK				
hBAP-1(483-729)	KFKIDDDQRRTHNYDEFICTFISMLAQEGMLANLVEQNISVRRRQGVSIG									
mBAP-1(518-(del)-718)	-----	KFKIDDDQRRTHNYDEFICTFISMLAQEGMLANLVEQNISVRRRQGVSIG								
mBAP-1(581-720)	KFKIDDDQRRTHNYDEFICTFISMLAQEGMLANLVEQNISVRRRQGVSIG									
mBAP-1(596-721)	KFKIDDDQRRTHNYDEFICTFISMLAQEGMLANLVEQNISVRRRQGVSIG									
hBAP-1(483-729)	RLHKQRKPD	RRKRS	RPYK	AKRQ	*					
mBAP-1(518-(del)-718)	RLHKQRKPD	RRR	-----	-----	-----	ms	g	r		
mBAP-1(581-720)	RLHKQRKPD	RRKR	-----	-----	-----	l	s	g	r	
mBAP-1(596-721)	RLHKQRKPD	RRKRS	-----	-----	-----	e	r	p	i	d

Fig. 2B

6 / 1 1

FIGURE 3B

CTG	GAT	GGG	CTG	AAG	GTC	TAC	CCC	ATT	GAC	CAT	GGG	CCC	TGG	GGG	630
Leu	Asp	Gly	Leu	Lys	Val	Tyr	Pro	Ile	Asp	His	Gly	Pro	Trp	Gly	
		185					190					195			
GAG	GAC	GAG	GAG	TGG	ACA	GAC	AAG	GCC	CGG	CGG	GTC	ATC	ATG	GAG	675
Glu	Asp	Glu	Glu	Trp	Thr	Asp	Lys	Ala	Arg	Arg	Val	Ile	Met	Glu	
		200					205					210			
CGT	ATC	GGC	CTC	GCC	ACT	GCA	GGG	GAG	CCC	TAC	CAC	GAC	ATC	CGC	720
Arg	Ile	Gly	Leu	Ala	Thr	Ala	Gly	Glu	Pro	Tyr	His	Asp	Ile	Arg	
		215					220					225			
TTC	AAC	CTG	ATG	GCA	GTG	GTG	CCC	GAC	CGC	AGG	ATC	AAG	TAT	GAG	765
Phe	Asn	Leu	Met	Ala	Val	Val	Pro	Asp	Arg	Arg	Ile	Lys	Tyr	Glu	
		230					235					240			
GCC	AGG	CTG	CAT	GTG	CTG	AAG	GTG	AAC	CGT	CAG	ACA	GTA	CTA	GAG	810
Ala	Arg	Leu	His	Val	Leu	Lys	Val	Asn	Arg	Gln	Thr	Val	Leu	Glu	
		245					250					255			
GCT	CTG	CAG	CAG	CTG	ATA	AGA	GTA	ACA	CAG	CCA	GAG	CTG	ATT	CAG	855
Ala	Leu	Gln	Gln	Leu	Ile	Arg	Val	Thr	Gln	Pro	Glu	Leu	Ile	Gln	
		260					265					270			
ACC	CAC	AAG	TCT	CAA	GAG	TCA	CAG	CTG	CCT	GAG	GAG	TCC	AAG	TCA	900
Thr	His	Lys	Ser	Gln	Glu	Ser	Gln	Leu	Pro	Glu	Glu	Ser	Lys	Ser	
		275					280					285			
GCC	AGC	AAC	AAG	TCC	CCG	CTG	GTG	CTG	GAA	GCA	AAC	AGG	GCC	CCT	945
Ala	Ser	Asn	Lys	Ser	Pro	Leu	Val	Leu	Glu	Ala	Asn	Arg	Ala	Pro	
		290					295					300			
GCA	GCC	TCT	GAG	GGC	AAC	CAC	ACA	GAT	GGT	GCA	GAG	GAG	GCG	GCT	990
Ala	Ala	Ser	Glu	Gly	Asn	His	Thr	Asp	Gly	Ala	Glu	Glu	Ala	Ala	
		305					310					315			
GGT	TCA	TGC	GCA	CAA	GCC	CCA	TCC	CAC	AGC	CCT	CCC	AAC	AAA	CCC	1035
Gly	Ser	Cys	Ala	Gln	Ala	Pro	Ser	His	Ser	Pro	Pro	Asn	Lys	Pro	
		320					325					330			
AAG	CTA	GTG	GTG	AAG	CCT	CCA	GGC	AGC	AGC	CTC	AAT	GGG	GTT	CAC	1080
Lys	Leu	Val	Val	Lys	Pro	Pro	Gly	Ser	Ser	Leu	Asn	Gly	Val	His	
		335					340					345			
CCC	AAC	CCC	ACT	CCC	ATT	GTC	CAG	CGG	CTG	CCG	GCC	TTT	CTA	GAC	1125
Pro	Asn	Pro	Thr	Pro	Ile	Val	Gln	Arg	Leu	Pro	Ala	Phe	Leu	Asp	
		350					355					360			
AAT	CAC	AAT	TAT	GCC	AAG	TCC	CCC	ATG	CAG	GAG	GAA	GAA	GAC	CTG	1170
Asn	His	Asn	Tyr	Ala	Lys	Ser	Pro	Met	Gln	Glu	Glu	Glu	Asp	Leu	
		365					370					375			

7 / 1 1

FIGURE 3C

GCG	GCA	GGT	GTG	GGC	CGC	AGC	CGA	GTT	CCA	GTC	CGC	CCA	CCC	CAG	1215
Ala	Ala	Gly	Val	Gly	Arg	Ser	Arg	Val	Pro	Val	Arg	Pro	Pro	Gln	
		380					385					390			
CAG	TAC	TCA	<u>GAT</u>	<u>GAT</u>	<u>GAG</u>	<u>GAT</u>	<u>GAC</u>	<u>TAT</u>	<u>GAG</u>	<u>GAT</u>	<u>GAC</u>	<u>GAG</u>	<u>GAG</u>	<u>GAT</u>	1260
Gln	Tyr	Ser	<u>Asp</u>	<u>Asp</u>	<u>Glu</u>	<u>Asp</u>	<u>Asp</u>	<u>Tyr</u>	<u>Glu</u>	<u>Asp</u>	<u>Asp</u>	<u>Glu</u>	<u>Glu</u>	<u>Asp</u>	
		395					400					405			
<u>GAC</u>	GTG	CAG	AAC	ACC	AAC	TCT	GCC	CTT	AGG	TAT	AAG	GGG	AAG	GGA	1305
<u>Asp</u>	Val	Gln	Asn	Thr	Asn	Ser	Ala	Leu	Arg	Tyr	Lys	Gly	Lys	Gly	
		410					415					420			
ACA	GGG	AAG	CCA	GGG	GCA	TTG	AGC	GGT	TCT	GCT	GAT	GGG	CAA	CTG	1350
Thr	Gly	Lys	Pro	Gly	Ala	Leu	Ser	Gly	Ser	Ala	Asp	Gly	Gln	Leu	
		425					430					435			
TCA	GTG	CTG	CAG	CCC	AAC	ACC	ATC	AAC	GTC	TTG	GCT	GAG	AAG	CTC	1395
Ser	Val	Leu	Gln	Pro	Asn	Thr	Ile	Asn	Val	Leu	Ala	Glu	Lys	Leu	
		440					445					450			
AAA	GAG	TCC	CAG	AAG	GAC	CTC	TCA	ATT	CCT	CTG	TCC	ATC	AAG	ACT	1440
Lys	Glu	Ser	Gln	Lys	Asp	Leu	Ser	Ile	Pro	Leu	Ser	Ile	Lys	Thr	
		455					460					465			
AGC	AGC	GGG	GCT	GGG	AGT	CCG	GCT	GTG	GCA	GTG	CCC	ACA	CAC	TCG	1485
Ser	Ser	Gly	Ala	Gly	Ser	Pro	Ala	Val	Ala	Val	Pro	Thr	His	Ser	
		470					475					480			
<u>CAG</u>	CCC	TCA	CCC	ACC	CCC	AGC	AAT	GAG	AGT	ACA	GAC	ACG	GCC	TCT	1530
<u>Gln</u>	Pro	Ser	Pro	Thr	Pro	Ser	Asn	Glu	Ser	Thr	Asp	Thr	Ala	Ser	
		485					490					495			
GAG	ATC	GGC	AGT	GCT	TTC	AAC	TCG	CCA	CTG	CGC	TCG	CCT	ATC	CGC	1575
Glu	Ile	Gly	Ser	Ala	Phe	Asn	Ser	Pro	Leu	Arg	Ser	Pro	Ile	Arg	
		500					505					510			
TCA	GCC	AAC	CCG	ACG	CGG	CCC	TCC	AGC	CCT	GTC	ACC	TCC	CAC	ATC	1620
Ser	Ala	Asn	Pro	Thr	Arg	Pro	Ser	Ser	Pro	Val	Thr	Ser	His	Ile	
		515					520					525			
TCC	AAG	GTG	CTT	TTT	GGA	GAG	GAT	GAC	AGC	CTG	CTG	CGT	GTT	GAC	1665
Ser	Lys	Val	Leu	Phe	Gly	Glu	Asp	Asp	Ser	Leu	Leu	Arg	Val	Asp	
		530					535					540			
TGC	ATA	CGC	TAC	AAC	CGT	GCT	GTC	CGT	GAT	CTG	GGT	CCT	GTC	ATC	1710
Cys	Ile	Arg	Tyr	Asn	Arg	Ala	Val	Arg	Asp	Leu	Gly	Pro	Val	Ile	
		545					550					555			
AGC	ACA	GGC	CTG	CTG	CAC	CTG	GCT	GAG	GAT	GGG	GTG	CTG	AGT	CCC	1755
Ser	Thr	Gly	Leu	Leu	His	Leu	Ala	Glu	Asp	Gly	Val	Leu	Ser	Pro	
		560					565					570			

8 / 1 1

FIGURE 3D

CTG GCG CTG <u>ACA</u> GAG GGT GGG AAG GGT TCC TCG CCC TCC ATC AGA	1800
Leu Ala Leu <u>Thr</u> Glu Gly Gly Lys Gly Ser Ser Pro Ser Ile Arg	
575 580 585	
CCA ATC CAA GGC AGC CAG GGG TCC AGC AGC CCA GTG GAG AAG GAG	1845
Pro Ile Gln Gly Ser Gln Gly Ser Ser Ser Pro Val Glu Lys Glu	
590 595 600	
GTC GTG GAA GCC ACG GAC AGC AGA GAG AAG ACG GGG ATG GTG AGG	1890
Val Val Glu Ala Thr Asp Ser Arg Glu Lys Thr Gly Met Val Arg	
605 610 615	
CCT GGC GAG CCC TTG AGT GGG GAG AAA TAC TCA CCC AAG GAG CTG	1935
Pro Gly Glu Pro Leu Ser Gly Glu Lys Tyr Ser Pro Lys Glu Leu	
620 625 630	
CTG GCA CTG CTG AAG TGT GTG GAG GCT GAG ATT GCA AAC TAT GAG	1980
Leu Ala Leu Leu Lys Cys Val Glu Ala Glu Ile Ala Asn Tyr Glu	
635 640 645	
GCG TGC CTC AAG GAG GAG GTA GAG AAG AGG AAG AAG TTC AAG ATT	2025
Ala Cys Leu Lys Glu Glu Val Glu <u>Lys Arg Lys Lys Phe Lys</u> Ile	
650 655 660	
GAT GAC CAG AGA AGG ACC CAC AAC TAC GAT GAG TTC ATC TGC ACC	2070
Asp Asp Gln Arg Arg Thr His Asn Tyr Asp Glu Phe Ile Cys Thr	
665 670 675	
TTT ATC TCC ATG CTG GCT CAG GAA GGC ATG CTG GCC AAC CTA GTG	2115
Phe Ile Ser Met Leu Ala Gln Glu Gly Met Leu Ala Asn Leu Val	
680 685 690	
GAG CAG AAC ATC TCC GTG CGG CGG CGC CAA GGG GTC AGC ATC GGC	2160
Glu Gln Asn Ile Ser Val Arg Arg Arg Gln Gly Val Ser Ile Gly	
695 700 705	
CGG CTC CAC AAG CAG CGG AAG CCT GAC CGG CGG AAA CGC TCT CGC	2205
Arg Leu His Lys Gln Arg Lys Pro Asp <u>Arg Arg Lys Arg Ser Arg</u>	
710 715 720	
CCC TAC AAG GCC AAG CGC CAG TGAGGACTGC TGGCCCTGAC TCTGCAGCCC	2256
Pro Tyr Lys Ala Lys Arg Gln	
725	
ACTCTTGCCG TGTGGCCCTC ACCAGGGTCC TTCCCTGCCC CACTTCCCCT	2306
TTTCCCAGTA TTACTGAATA GTCCCAGCTG GAGAGTCCAG GCCCTGGGAA	2356
TGGGAGGAAC CAGGCCACAT TCCTTCCATC GTGCCCTGAG GCCTGACACG	2406
GCAGATCAGC CCCATAGTGC TCAGGAGGCA GCATCTGGAG TTGGGGCACA	2456
GCGAGGTACT GCAGCTTCCT CCACAGCCGG CTGTGGAGCA GCAGGACCTG	2506

9 / 1 1

FIGURE 3E

GCCCTTCTGC CTGGGCAGCA GAATATATAT TTTACCTATC AGAGACATCT	2556
ATTTTTCTGG GCTCCAACCC AACATGCCAC CATGTTGACA TAAGTTCCTA	2606
CCTGACTATG CTTTCTCTCC TAGGAGCTGT CCTGGTGGGC CCAGGTCCTT	2656
GTATCATCCA CGGTCCCAAC TACAGGGTCC TAGCTGGGGG CCTGGGTGGG	2706
CCCTGGGCTC TGGGCCCTGC TGCTCTAGCC CCAGCCACCA GCCTGTCCCT	2756
GTTGTAAGGA AGCCAGGTCT TCTCTCTTCA TTCCTCTTAG GAGAGTGCCA	2806
AACTCAGGGA CCCAGCACTG GGCTGGGTTG GGAGTAGGGT GTCCCAGTGG	2856
GGTTGGGGTG AGCAGGCTGC TGGGATCCCA TGGCCTGAGC AGAGCATGTG	2906
GGAAGTGTTC AGTGGCCTGT GAACTGTCTT CCTTGTTCTA GCCAGGCTGT	2956
TCAAGACTGC TCTCCATAGC AAGGTTCTAG GGCTCTTCGC CTTCAGTGTT	3006
GTGGCCCTAG CTATGGGCCT AAATTGGGCT CTAGGTCTCT GTCCCTGGCG	3056
CTTGAGGCTC AGAAGAGCCT CTGTCCAGCC CCTCAGTATT ACCATGTCTC	3106
CCTCTCAGGG GTAGCAGAGA CAGGGTTGCT TATAGGAAGC TGGCACCCT	3156
CAGCTCTTCC TGCTACTCCA GTTTCCTCAG CCTCTGCAAG GCACTCAGGG	3206
TGGGGGACAG CAGGATCAAG ACAACCCGTT GGAGCCCCTG TGTTCCAGAG	3256
GACCTGATGC CAAGGGGTAA TGGGCCCAGC AGTGCCTCTG GAGCCCAGGC	3306
CCCAACACAG CCCCATGGCC TCTCCAGATG GCTTTGAAAA GGTGATCCAA	3356
CAGGCCCCTT TATCTGTACA TAGTGACTGA GTGGGGGGTG CTGGCAAGTG	3406
TGGCACTCCT CTGGGCTGAG CACAGCTTGA CCCCTCTAGC CCCTGTAAAA	3456
CTGGATCAAT GAATGAATAA AACTCTCCTA AGATCTCCTG AGAAAAAAAA	3506
AAAAAAAAAG G	3517

10 / 1 1

FIG. 3F

BAP1	1	-----MNRGNLELESDPGLFTLLVEDPQVKG-VQVREIYDLQS--	37
CREEL-cos11.7	1	MCKKIMTDAGSMCLIESDPGVFTENLRQFQVDC-LQVEELYSLEDD--	44
UCH-DROME	1	-----MLTNTPLLESHPEVLTXYIKKLAVSPANSVTDVIGLED-D	38
YUH1	1	-----MSGENRAVVPFIESHPFVFTNFARKLGLXHNENAYFDIYSLTEPE	43
UCH-L3	1	-----MEGQRNLPLEAMPFVTHQFLKQLGLFPHNQFVVDVYGMDF-E	48
UCH-L1	1	-----MQLKPHFIDFPEMLNKKVLSALGVAGQWRFPVDVLCLEEE-E	37
BAP1	38	KC---QGPPVYGFIFLPKXIEERASARKVSTLVDDTSVIDDDIVNMMF	81
CREEL-cos11.7	45	-DKAMTRPTTGLIFLPKWRQODE-----TTGIF-----SDKQHF	78
UCH-DROME	39	TLEWIFRPVKAIFLLFPCHETTEKXR--TEEDRIKEVEEQEPEDLF	83
YUH1	44	LLAPLPFPVKAIVLLFPFTHEDRKSST--SQQIT-----SSYDVI	88
UCH-L3	41	LLSNVFPFVCAVLLLPFITEKIEVFR--TEEEKIKSQGQDVTSSVY	85
UCH-L1	38	SLGSPAPACALLLLFPPLTAQHNF--KKQIEELK--GQEVSPKVF	88
BAP1	82	FAGQLIPNSCATHALLSVLLNCSS--VD-LGPTLSRMKDPFTKGFSP-	124
CREEL-cos11.7	79	FAGOTIQNACATQALINLNNVED-TDVKLGNILNQYKEFAIDLDP-	123
UCH-DROME	84	YMRQPTENACGTVALIHSVAHKE-VDIDRG-VLKDPLEKTASLSP-	127
YUH1	81	WFKQSVKMACGLYALLESLSNNQS--LLEPGSDLDNFKLSQSDTSSS	125
UCH-L3	86	FMKQTIENACGTICGLIHAIAHKKDKHFESESTLKKFLZESVMSFP-	131
UCH-L1	81	FMRQITIGNSCGTICGLIHAIAHKKDKHFESESTLKKFLZESVMSFP-	126
BAP1	125	ESKGYAIGNAPELAKAENSHARPEPRHLPEKQNGLSAVRTMEAFHFV	171
CREEL-cos11.7	124	NTDGECLSHSEIIRTVENSPSR-----QTLFELDIXGGESEDNHFV	165
UCH-DROME	128	EERGRALEXDEXFTADREALAQ-----EGQ--THAANHEKVIHHTI	166
YUH1	126	KNRTPDDVTTDQFVLNVIKENVQT--FSTGQSEAPEATADTNLHTI	168
UCH-L3	132	EEARYLENYDAIRVTHETSAN-----EGQTEAPSIDEKVDLHTI	171
UCH-L1	127	EDRAKCFEKNEAIQAABDAVAQ-----EGQ--CR-VDDKVNTHI	163
BAP1	172	SYVPITGRLELDGL-KVYFIDEG---PNGEDEENTDKARRVIMERI	214
CREEL-cos11.7	166	TYVPIGNKVVYELDGL-RELPLEVA---EPQKEQDNWIAIKPVIQQRH	208
UCH-DROME	167	ALVVKEGTLYELDGR-KSPFIKSG---PTSEE-TFVKDAAKVCKEPM	208
YUH1	169	TYVEENGQIFELDGRNLSGLYLGKSDPTATD-LIEQELVRVRVASY	214
UCH-L3	172	ALVEVDGELYELDGR-KPFPINHG---ETSDE-TLLEDAIEVCKKPM	213
UCH-L1	164	LFNNVDGELYELDGR-MPFPVNEG---ASSED-TLLKDAAKVCREPT	205
BAP1	215	GLATAGEPYHDIRFNLMAVVPDARIKYEARLEVLKVNROTVLEALQQ	261
CREEL-cos11.7	209	QKYSGE-----ITFNLMAVFPNRKQKLQEMMENLIQANENNLEEQI	251
UCH-DROME	209	ARD-PHE---VRFVLAALTAQQ	227
YUH1	215	MEN-ANED-VLNFAMLGLOPHWE	236
UCH-L3	214	ERD-PDE---LRFNAIALSAA	238
UCH-L1	206	ERE-QGE---VRFSAVALCKAA	223
BAP1	638	KELLALLKCVAEAEIANYTEACLKEEVZKRRKKFKID	633
CREEL-cos11.7	238	LIQANENNELIEQIADLNKAIADEDIYKMEHYRKE	271
BAP1	634	DQRRRTBNYDEFFICTYISM LAQEGMLANLVLEQNIS	667
CREEL-cos11.7	272	NNRRRTBNYTFVFIELMKILAKEGKLVGLVDNAYQ	385
BAP1	668	VRRRQGVSIQRLHKKQKKPDRRKRSRPFYKAKRQ	729
CREEL-cos11.7	386	AAK-ZK-S--KLN TDITKLELKKRKQ	326

11 / 1 1

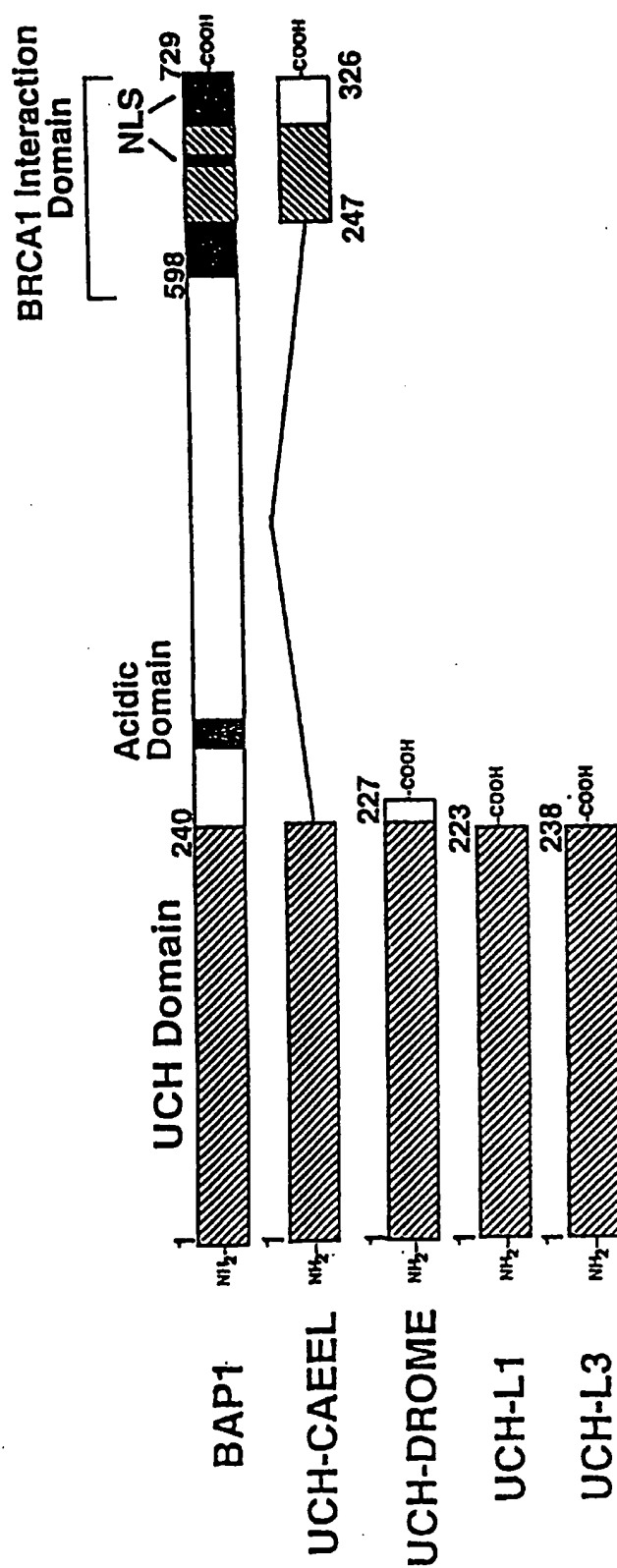


FIG. 3G

INTERNATIONAL SEARCH REPORT

 International application No.
PCT/US97/13684

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) : G01N 33/574; C07K 16/30; C07H 17/00

US CL : Please See Extra Sheet

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 536/23.1, 24.31; 530/387.7, 387.2, 350; 435/6, 7.23, 325; 436/501, 64

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS; DIALOG: file biochem; GENBANK, EMBL, EST-STS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WEBER et al. Familial Breast Cancer - Approaching the Isolation of a Susceptibility Gene. Cancer. 01 August 1994, Vol. 74, No. 3, pages 1013-1020.	1-22
A	FORD et al. The Genetics of Breast and Ovarian Cancer. British Journal of Cancer. 1995, Vol. 72, pages 805-812.	1-22
Y, P	Database DDBJ/EMBL/Genbank, Accession No. D87462, NOMURA et al. 'Prediction of the coding sequences of unidentified human genes,' Direct Submission 27 August 1996.	1-22
Y	Database EST-STS, Accession No. N77549, HILLIER et al. 'The Washu-Merck EST Project,' Direct Submission 02 April 1996.	1-22

☒ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:	*T	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
A document defining the general state of the art which is not considered to be of particular relevance	*X	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
B earlier document published on or after the international filing date	*Y	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
L document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	*A*	document member of the same patent family
O document referring to an oral disclosure, use, exhibition or other means		
P document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 28 SEPTEMBER 1997	Date of mailing of the international search report 13 JAN 1998
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer TOMAS SCHNEIDER Telephone No. (703) 305-0196

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US97/13684

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	Database EST-STS-Three, Accession No. W35227, HILLIER et al. 'The WashU-Merck EST Project,' Direct Submission 16 May 1996.	1-22

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US97/13684

A. CLASSIFICATION OF SUBJECT MATTER:

US CL :

536/23.1, 24.31; 530/387.7, 387.2, 350; 435/6, 7.23, 325; 436/501, 64

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☒ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☒ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.